

1-1-2011

The Role of Feedback in the Assimilation of Information in Prediction Markets

Richard Donald Jolly
Portland State University

Let us know how access to this document benefits you.

Follow this and additional works at: http://pdxscholar.library.pdx.edu/open_access_etds

Recommended Citation

Jolly, Richard Donald, "The Role of Feedback in the Assimilation of Information in Prediction Markets" (2011). *Dissertations and Theses*. Paper 468.

[10.15760/etd.468](https://pdxscholar.library.pdx.edu/open_access_etds/10.15760/etd.468)

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

The Role of Feedback in the Assimilation of Information in Prediction
Markets

by

Richard Donald Jolly

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy
in
Systems Science: Business Administration

Dissertation Committee:
Wayne Wakeland, Chair
Barry Anderson
Steven Bleiler
James Woods
Martin Zwick

Portland State University
© 2011

Abstract

Leveraging the knowledge of an organization is an ongoing challenge that has given rise to the field of knowledge management. Yet, despite spending enormous sums of organizational resources on Information Technology (IT) systems, executives recognize there is much more knowledge to harvest. Prediction markets are emerging as one tool to help extract this tacit knowledge and make it operational. Yet, prediction markets, like other markets, are subject to pathologies (e.g., bubbles and crashes) which compromise their accuracy and may discourage organizational use.

The techniques of experimental economics were used to study the characteristics of prediction markets. Empirical data was gathered from an on-line asynchronous prediction market. Participants allocated tickets based on private information and, depending on the market type, public information indicative of how prior participants had allocated their tickets. The experimental design featured three levels of feedback (no-feedback, percentages of total allocated tickets and frequency of total allocated tickets) presented to the participants.

The research supported the hypothesis that information assimilation in feedback markets is composed of two mechanisms - information collection and aggregation. These are defined as:

- Collection - The compilation of dispersed information - individuals using their own private information make judgments and act accordingly in the market.

- Aggregation - The market's judgment on the implications of this gathered information - an inductive process. This effect comes from participants integrating public information with their private information in their decision process.

Information collection was studied in isolation in no feedback markets and the hypothesis that markets outperform the average of their participants was supported. The hypothesis that with the addition of feedback, the process of aggregation would be present was also supported. Aggregation was shown to create agreement in markets (as measured by entropy) and drive market results closer to correct values (the known probabilities). However, the research also supported the hypothesis that aggregation can lead to information mirages, creating a market bubble.

The research showed that the presence and type of feedback can be used to modulate market performance. Adding feedback, or more informative feedback, increased the market's precision at the expense of accuracy. The research supported the hypotheses that these changes were due to the inductive aggregation process which creates agreement (increasing precision), but also occasionally generates information mirages (which reduces accuracy).

The way individual participants use information to make allocations was characterized. In feedback markets the fit of participants' responses to various decision models demonstrated great variety. The decision models ranged from little use of information (e.g., MaxiMin), use of only private information (e.g., allocation in proportion to

probabilities), use of only public information (e.g., allocating in proportion to public distributions) and integration of public and private information. Analysis of all feedback market responses using multivariate regression also supported the hypothesis that public and private information were being integrated by some participants. The subtle information integration results are in contrast to the distinct differences seen in markets with varying levels of feedback. This illustrates that the differences in market performance with feedback are an emergent phenomenon (i.e., one that could not be predicted by analyzing the behavior of individuals in different market situations).

The results of this study have increased our collective knowledge of market operation and have revealed methods that organizations can use in the construction and analysis of prediction markets. In some situations markets without feedback may be a preferred option. The research supports the hypothesis that information aggregation in feedback markets can be simultaneously responsible for beneficial information processing as well as harmful information mirage induced bubbles. In fact, a market subject to mirage prone data resembles a Prisoner's Dilemma where individual rationality results in collective irrationality.

Dedication

This work is dedicated to my past, my present and my future ...

The beloved memory of my parents, Charles and Ruth

The love and devotion of my wonderful wife, Rialina

The hope of my children, Elena and Elise

Acknowledgements

I wish every Ph.D. candidate might have a committee like mine . Systems Science is inherently interdisciplinary, and I've been able to assemble a committee from the disciplines of Systems Science, Economics, Psychology, Mathematics and Business Administration. I deeply thank all my committee members Dr. Barry Anderson, Dr. Steven Bleiler, Dr. James Woods and Dr. Martin Zwick for their constructive comments, helpful suggestions, enthusiasm and encouragement! A special thanks is deserved to Dr. Wayne Wakeland, my committee chair, for being so generous with his time and so helpful in idea brainstorming. I'd also like to thank my advisor, Dr. John Settle, who was on my committee for most of its existence - leaving only because of his retirement from the University.

I was working for Intel for most of my Ph.D. candidacy, and I thank them for their support. A special thanks to Jay Hopman from Intel who shared much of his knowledge about prediction markets with me.

I'd like to thank the Dean of the School of Business, Dr. Scott Dawson, for early discussions which led me into the Ph.D. program. Dr. Todd Bodner, in addition to serving on my Comprehensive exam committee, helped with several questions about the experimental design. Dr. Robert Sinclair, who also served on my Comprehensive exam committee, provided valuable guidance and support.

I'd like to thank Dr. Charles Plott for sharing so much of his knowledge on a visit to Caltech with Jay Hopman and myself in support of Intel's prediction markets. I'd like to sincerely thank Ken Luke who developed the survey implementation software - a very large and complicated task.

A Ph.D. program lasts many years and I've benefited from so many conversations with faculty, students and University staff that I can't even recall, much less cite them. Still, they were all instrumental in the process and I thank each and every one.

No Ph.D. can be completed without the support of the candidate's family. I'd like to thank my mother-in-law, Nadiia, for her help with the children over this period. My children, Elena and Elise, have reminded me about pure, unencumbered, curiosity. Finally, and most importantly, I thank my wife Rialina, for all her love and support over the term of this work.

Table of Contents

Abstract.....	i
Dedication.....	iv
Acknowledgments	v
List of Tables.....	x
List of Figures.....	xii
1 Motivations and Research Questions.....	1
1.1 Motivations.....	1
1.1.1 Organization Knowledge Management	1
1.1.2 Issues with Sharing Information	2
1.1.3 Wisdom of Crowds	2
1.1.4 Prediction Markets and Organizational Use	6
1.1.5 Accuracy of Prediction markets.....	7
1.1.6 Limitations of Prediction Markets	7
1.1.7 The Role of Feedback in Markets.....	8
1.1.8 Motivations of the Research	8
1.2 Research Questions	9
1.2.1 What are the Mechanisms Involved in the Information Assimilation Process?	9
1.2.2 How does the Presence and Type of Feedback impact Market Performance?	9
1.2.3 How do Individuals use Information in the Allocation Process?	9
2 Literature Review, Synthesis and Experimental Hypotheses	11
2.1 Market Background.....	11
2.1.1 Market Types	11
2.1.2 Information Assimilation Capability of Markets.....	12
2.1.3 Prediction Markets	18
2.1.4 Experimental Prediction Markets	19
2.2 Individual Decision Processes.....	23
2.2.1 Overview.....	23
2.2.2 Judgment - Assessment of Probability.....	24
2.2.3 Decisions - Allocation of Resources.....	29
2.3 Interactions in Markets.....	33
2.3.1 Rational Markets with Irrational Agents.....	34
2.3.2 Feedback	35
2.3.3 Systems Theories of markets	40
2.4 Decisions Frameworks in Economics	42
3 Experimental Hypotheses and Methods.....	44
3.1 Synthesis and Experimental Hypotheses	44
3.1.1 Research Framework and Market Measurements.....	44
3.1.2 What are the Mechanisms Involved in the Information Assimilation Process?	50
3.1.3 How does the Presence and Type of Feedback impact Market Performance?	53

3.1.4	How do Individuals use Information in the Allocation Process?	55
3.1.5	Summary of Research Hypotheses	57
3.2	Experiment Design	59
3.2.1	Simplified Experimental Prediction Market	60
3.2.2	Random Draw Experiment	63
3.2.3	Mirage Prone Draw Experiment	63
3.3	Survey Implementation	66
3.3.1	Staging in Random Draw Experiment	69
3.3.2	Staging in Mirage Prone Draw Experiment.....	70
3.4	Research Analysis Overview and Tools.....	73
3.4.1	Evaluating Draw Sets.....	73
3.4.2	Understanding Individual Allocations	75
3.4.3	Inferential Tools and Considerations	81
3.5	Potential Limitations of Experimental Methods	82
3.5.1	Sequencing Participants	82
3.5.2	Learning	83
4	Experimental Results	84
4.1	General Considerations	84
4.1.1	Data Collection	84
4.1.2	Demographics	85
4.1.3	Careless Responses	86
4.1.4	Screening.....	90
4.2	Market Performance with Random Draws.....	91
4.2.1	Information Collection in Markets with No-feedback.....	91
4.2.2	Information Aggregation in Markets with Feedback.....	93
4.2.3	Mirage Tendencies in Feedback Markets	108
4.3	Market Performance under Mirage Prone Conditions	113
4.3.1	Market's Performance with Mirage Prone Draw Sets.....	113
4.3.2	Evaluation for Information Mirages	119
4.4	Individual Allocations.....	128
4.4.1	Rounding and Position Bias.....	128
4.4.2	No-feedback Markets: Allocations vs. Probability	130
4.4.3	Feedback Markets: Influence of Public Data on Allocations	131
4.4.4	Categorizing responses into styles	137
5	Discussion, Conclusions and Next Steps	142
5.1	Conclusions	142
5.1.1	What are the Mechanisms Involved in the Information Assimilation Process? 142	
5.1.2	How does the Presence and Type of Feedback Impact Market Performance? 150	
5.1.3	How do Individuals use Information in the Allocation Process?	153
5.2	Simulating Markets	162
5.2.1	Simulation Goals.....	162
5.2.2	Characteristics of Interest	163
5.2.3	Simulator Architecture.....	165

5.2.4	Simulator Calibration and Validation	167
5.2.5	Initial Simulation Results.....	171
5.3	Considerations for Prediction Market Designers	173
5.3.1	Choosing a Feedback Type.....	173
5.3.2	Market Size (Number of Participants)	177
5.3.3	Guiding Participants to Improve Performance	178
5.3.4	Screening.....	179
5.3.5	Synchronous or Asynchronous Markets	180
5.3.6	Evaluating Prediction Market Results	180
5.4	Future Research.....	181
5.4.1	Simulation	181
5.4.2	Parimutuel Markets	183
5.4.3	Hybrid Markets	186
5.4.4	Treatments to Improve Individual's Performance.....	187
5.4.5	Correlation of Allocation Styles to Psychological Characteristics	187
5.4.6	Impact of Final Digit Bias.....	187
5.5	Summary	187
6	References	191
	Appendix - Survey Details.....	198

List of Tables

Table 1. Summary of research questions.	10
Table 2. Summary of results from Roust & Plott Experiments (2005)	22
Table 3. Bayesian posterior probabilities of private information	25
Table 4. Illustration of choices from Allais paradox	31
Table 5. Illustration of two stage evaluation of allocations	47
Table 6. Contingency table to test differences in decisiveness	48
Table 7. Contingency table to test differences in correctness.....	48
Table 8. Summary of market measurement instruments.	50
Table 9. Summary of research hypotheses for the first research question.....	53
Table 10. Market Structure factors of Interest	54
Table 11. Summary of research hypotheses for the second research question.....	55
Table 12. Summary of research hypotheses for the third research question	57
Table 13. Summary of research questions	58
Table 14. Overview of experimental design	59
Table 15. Experimental Draw Table for Mirage Prone Draw Experiment.....	64
Table 16. Statistics of the draw table compared to the Urn.	64
Table 17. Mirage prone draw set and AIA posterior probability.....	65
Table 18. Accumulated allocation	66
Table 19. Staging for random draws experiment.....	70
Table 20. Illustration of the staging method for the mirage prone experiment	72
Table 21. Illustration of the staging method for the mirage prone experiment	73
Table 22. Bayesian probabilities given the draw sets shown.....	77
Table 23. Categorization of responses in IP market to theoretical models.....	77
Table 24. Survey Data Collection Summary	85
Table 25. Summary of Demographic Responses.....	86
Table 26. IP market with spurious input from Participant 5.....	87
Table 27. All allocations from participant highlighted in Table 26.....	89
Table 28. Predicted responses to ECE draw for each allocation method.	90
Table 29. Illustration of individual versus market results.....	92
Table 30. Summary of decisiveness and correctness.....	93
Table 31. No-feedback and feedback market results as counts	95
Table 32. No-feedback and feedback market results as percentages.....	96
Table 33. Decisiveness and correctness for feedback and no-feedback markets	97
Table 34. Market results for frequency and percent markets as counts.....	98
Table 35. Market results for frequency and percent markets as percentage.....	99
Table 36. Decisiveness and correctness for the two types of feedback markets studied	100
Table 37. Mean value of Wurtz distance	102
Table 38. Mean entropy by stage.....	103
Table 39. Characteristics of the entropy distribution.....	105
Table 40. Characteristics of the entropy distribution.....	106
Table 41. Frequency market which exhibits an information mirage.	109
Table 42. Histogram of states for draw set shown in Table 41.....	110
Table 43. Percentage feedback market which exhibits an information mirage.	111

Table 44. Histogram of the states in the draw set in Table 43.....	111
Table 45. Performance of no-feedback and feedback markets	114
Table 46. Performance of no-feedback and feedback markets	115
Table 47. Decisiveness and correctness.....	116
Table 48. Comparison of decisiveness and correctness.....	117
Table 49. Performance of frequency and percent feedback markets	117
Table 50. Performance of frequency and percent feedback markets	118
Table 51. Comparison of decisiveness and correctness.....	119
Table 52. Mean Wurtz distance	120
Table 53. Mean entropy by stage.....	122
Table 54. Mirage prone draw set and the sum of draws by stage.....	124
Table 55. Mirage prone draw set and the sum of order weighted draws by stage.....	125
Table 56. Mean cumulative allocations by stage	126
Table 57. Mean cumulative allocations by stage	126
Table 58. Regression results for no-feedback markets	127
Table 59. Regression results for feedback markets	128
Table 60. Summary of regression analysis	134
Table 61. Summary of participant's allocations in no-feedback markets	134
Table 62. Multivariate Regression Results.	135
Table 63. Multivariate Regression Results.	136
Table 64. Linear relationships implied by the multivariate analysis.	136
Table 65. Descriptive results from categorization of participant's allocations	139
Table 66. Descriptive results from categorization of participant's allocations	139
Table 67. Grouping of styles into major categories.....	140
Table 68. Grouping of styles into major categories.....	140
Table 69. Categorization of participants' responses by use of information	141
Table 70. Comparison of decisiveness and correctness.....	146
Table 71. Summary of hypotheses and evaluations for the first research question -.....	149
Table 72. Summary of second research hypotheses and their evaluation.....	153
Table 73. Coefficients from Table 62	155
Table 74. Summary of hypotheses and their evaluation for the third research question	162
Table 75. Summary of factors which can be studied in a market simulator.	164
Table 76. Experimental vs. simulated decisiveness and Fisher's Exact test p-value.	169
Table 77. . Experimental vs. simulated correctness and Fisher's Exact test p-value	170
Table 78. Simulated market performance.....	172
Table 79. Comparison of no-feedback and feedback markets.....	175
Table 80. Percent correct by market type.	176
Table 81. Potential selection criteria for market structure.....	177
Table 82. Hypothesized impact of key factors.....	183
Table 83. Pros/Cons of parimutuel system	185

List of Figures

Figure 1. Steps in the ticket allocation process	23
Figure 2. Illustration of accuracy vs. precision	49
Figure 3. Experiment flow chart	60
Figure 4. Illustration of market operation	62
Figure 5. Survey Architecture	67
Figure 6. Actual implementation of the frequency feedback market	68
Figure 7. Mappings used to create alternate version of draw table	71
Figure 8. A well behaved draw table.	74
Figure 9. A non-well behaved draw table.	75
Figure 10. Qualitative comparison of basic allocation methods	76
Figure 11. Possible methods of using Bayesian probability	79
Figure 12. Summary of the allocation styles studied.	81
Figure 13. Summary of market performance	101
Figure 14. Histogram of final entropy distributions	107
Figure 15. Relationship between mean entropy and decisiveness	108
Figure 16. Wurtz distance between AIA and cumulative allocation	113
Figure 17. Wurtz distance between cumulative allocation	121
Figure 18. Entropy distributions for no-feedback and feedback markets	123
Figure 19. Histogram of final digit of allocation	129
Figure 20. Comparison of average ticket allocation by state to urn correct draw	130
Figure 21. Allocations vs. Bayesian probability - IP markets	131
Figure 22. Scatter plot of participant's allocation vs. the public cumulative	133
Figure 23. How participant's reported resolving conflicts	137
Figure 24. Summary of market performance	151
Figure 25. Allocations vs. Bayesian probability from no-feedback markets	154
Figure 26. Allocations vs. public data for 3 balls drawn.	157
Figure 27. Participant's allocation vs. public allocation	158
Figure 28. Simulated Prospect Theory Utility Curve	159
Figure 29. Market Simulator Architecture	167
Figure 30. Comparison of experimental and simulated decisiveness and correctness ...	171
Figure 31. Decisiveness and error rate observed in the random draws experiment	174
Figure 32. Decisiveness and correctness by stage (random draws experiment)	178

1 Motivations and Research Questions

The motivations and research questions for the present study are explained in this section.

1.1 Motivations

“If only HP knew what HP knows, we would be three times more productive”

Lew Platt, while CEO of Hewlett-Packard

Leveraging the knowledge of an organization is an ongoing challenge that has given rise to the field of knowledge management. Yet, despite spending enormous sums of organizational resources on Information Technology (IT) systems, executives recognize there is much more knowledge to harvest – as expressed by Lew Platt’s comment. The role and issues associated prediction markets as a tool to reveal latent knowledge within the organization are discussed below.

1.1.1 Organization Knowledge Management

Knowledge Management (KM) is a field within the area of Information Sciences which deals with the organizational use of information and knowledge. The field is of great practical value to organizational Information Technology (IT) groups – that is, the group that develops and deploys computer and technology systems for the organization’s employees. KM is discussed in numerous books and academic journals - cf. (Liebowitz, 1999), (Tiwana, 2002), (Davenport & Prusak, 1998), (Choo, 1998) and (Nissen, 2006).

1.1.2 Issues with Sharing Information

Studies have shown that knowledge management (KM) projects can encounter a variety of problems and challenges (c.f. Gupta, Iyer, & Aronson, 2000; Lin & Kwok, 2006; Sharp, 2003). Invariably, these problems are the result of unforeseen side effects manifested by the *system* being manipulated – namely the organization. Fundamentally, a KM system requires that members of the organization share their knowledge. This may simply mean making explicit knowledge accessible to others or transforming tacit knowledge into explicit knowledge for sharing. There are a number of reasons an individual may not be willing to share their knowledge (Ciborra & Patriotta, 1998) (Gilmour, 2003; Schutte & Snyman, 2006):

- Individuals may feel their *proprietary* knowledge is a competitive advantage versus their fellow employees – resulting in a prisoner’s dilemma (Jolly & Wakeland, 2008)
- They may fear loss of power or control
- They may fear ridicule or criticism
- They may not feel sharing is *fair*, as described by experimental economics research (Sigmund, Fehr, & Nowak, 2002).
- Organizational culture may not be conducive to knowledge sharing (Long & Fahey, 2000).

1.1.3 Wisdom of Crowds

The term *Wisdom of Crowds* was popularized by the best-selling book of the same title by James Surowiecki (2005), a business columnist for the *New Yorker* magazine. Surowiecki’s book is important for at least two reasons: 1) he does a credible job of collecting the most important ideas and research; and, 2) he presents the material in a very readable fashion, bringing the topic to a large group of business managers.

The thesis of the wisdom of crowds is that the diverse knowledge and expertise of a group can potentially be harnessed and applied to a range of problems. Further, the argument offers that, by virtue of the aggregated diversity, these groups can potentially outperform groups of experts. Several other books have been written on the general topic of crowd wisdom with similar discussions (Sunstein, 2006), (Benkler, 2006), (Tapscott & Williams, 2006), (Rheingold, 2002).

Surowiecki's main thesis is that there are four conditions that characterize wise crowds:

- Diversity of opinion
- Independence
- Decentralization (people are able to specialize and draw on local knowledge)
- An aggregation mechanism

These concepts are discussed below.

1.1.3.1 The Value of Diversity

Scott Page is professor and researcher at the University of Michigan. His book *The Difference* (Page, 2007) develops ideas around diversity in groups. One of the main propositions he develops is what he calls “the crowds beats the average law.” This asserts “Given a collection of diverse predictive models, the collective prediction is more accurate than the average individual predictions.” That is the *collective prediction error* is less than the *average individual error* (Hong & Page, 2004). Hong and Page develop this assertion via modeling and a mathematical derivation. Their derivation uses the set of assumptions listed below:

- Agents are intelligent (that is, they can find a marginally better solution)
- The problem is difficult (that is, no agent can always find the optimal solution)

- Diversity of agents (if the predictions differ then prediction diversity must be positive)
- The best agent is unique

Given these conditions, Hong and Page assert that groups of diverse problem solvers can outperform groups of high-ability problem solvers. So, the theorem asserts that both ability and diversity contribute to the wisdom of crowds.

1.1.3.2 The Power of Greater Numbers – and Independence

Nicolas de Condorcet, was a French philosopher and mathematician who developed what is today known as *Condorcet's Jury theorem* (Condorcet, 1785). The theorem states that for a group making a binary decision by majority rule then if each member's probability of making a correct choice is greater than $\frac{1}{2}$ the probability that the group's decision will be correct increases as the number of members of the group increases – approaching 1 as the group grows infinitely large.

In Condorcet's proof of the theorem, he assumed independence of the voters in order to eliminate dependent probabilities from the calculation. Ladha (1992) has studied Condorcet's theorem under the conditions of correlated votes. Ladha summarizes his findings as:

“for large groups, Condorcet's results would hold under fairly general conditions. For small groups, the conditions are severe. Finally, under reasonable assumptions, P_n , the probability that a majority selects the better alternative, would be inversely related to p , the average of the coefficients of correlation.”

In summary, given the assumption that each member's probability of being correct is greater than $\frac{1}{2}$, then:

- Greater numbers improves the result
- Greater independence improves the result

1.1.3.3 Aggregation Mechanisms

To extract the wisdom of the crowd a mechanism is needed to aggregate inputs. Chief among these mechanisms are polls and markets. Surowiecki opens his book by recounting the story of British scientist Francis Galton's description of a weight judging contest at the 1906 West of England Fat Stock and Poultry exhibition. For a small fee, fair participants could make a guess of the weight of an ox which was on display. The best guesses received prizes. Some 800 people entered the contest, a diverse mix of people attending the fair. After the contest Galton collected the tickets and analyzed them. He used the sample mean as a representation of the group's collective guess. The sample mean turned out to be 1,197 pounds. The correct weight of the ox was 1,198 pounds. This is an example of an incentivized poll.

Another prime example of an aggregation mechanism is a double auction market. Surowiecki illustrates the wisdom of crowds at work in a market with the stock market reaction immediately following the Challenger Space shuttle explosion. Within a few hours of the explosion, the stock price of Morton Thiokol, manufacturer of the solid rocket boosters which were the cause of the explosion, had dropped much lower than the stock of all the other space shuttle contractors (Maloney & Mulherin, 2003). However, there was no comment from NASA or any other authority on the cause of the accident for several days. The next day the NY Times noted "There are no clues to the cause of the accident." However, the stock market had already made a correct assessment.

1.1.4 Prediction Markets and Organizational Use

Prediction markets (also called predictive markets, information markets, decision markets, idea futures, event derivatives and virtual markets) are a type of market which can be used to harness the wisdom of crowds. A prediction market is a type of futures market designed to extract information about some type of future event. Typically, an instrument (e.g., a futures contract) is created whose final value is tied to the future event. The current market prices can be interpreted as predictions of the probability of the event or the expected value of the parameter. Example use of prediction markets by organizations include aids in demand forecasting (Hopman, 2007). Prediction markets are the topic of several current books – (e.g. Abramowicz, 2007; Hahn & Tetlock, 2006). Also, of note, analyst firms such as Forrester Research which do research, analysis and consultation for IT organizations have begun writing about and recommending prediction markets as an business tool (Young, 2008).

Prediction markets offer a financial incentive for organizational members to participate. This directly addresses some of the barriers to sharing mentioned above. Page also argues that incentives are important as they drive out less accurate predictions and more heavily weight accurate ones (as long as the predictors understand how accurate they might be).

Anonymity is another critical characteristic of prediction markets. This addresses some of the mentioned barriers to sharing as well as some of the group deliberation biases as described by Sunstein (2006).

Organizational prediction markets have been used for financial forecasts and schedules predictions. However, this tool has organization hurdles to overcome (H. Berg) including:

- Concern about self-fulfilling late forecasts (e.g., an employee makes a low prediction and proceeds to reduce his effort)
- Potential for negative consequences (e.g., if a prediction market predicts a project will miss a schedule, employees may reduce their work effort)
- Fit to the project process flow (e.g., data must be available when it can have an impact)
- Negative impressions around prediction markets (e.g., insider trading, gambling is bad, not appropriate for work, betting on failure, etc.)

1.1.5 Accuracy of Prediction markets

Berg et al. (2008) review the accuracy of Iowa Electronic Market's Political Market forecasts. They analyzed the results of 49 markets covering 41 elections in 13 countries. They not only compared the markets to the actual results, but also to leading opinion polls. Their results showed prediction markets exhibit excellent accuracy and no observed bias.

1.1.6 Limitations of Prediction Markets

However, prediction markets, like all markets, are subject to anomalies which can compromise their accuracy - market bubbles and crashes as prime examples.

Surowiecki chose the title of his book as a word-play off the title of another very famous book about crowd behavior, *Extraordinary Popular Delusions & the Madness of Crowds* by Charles Mackay (1932). Mackay's book, originally published in 1841, has several chapters on economic bubbles in history including the Dutch tulip mania of the early

seventeenth century and the South Sea Company bubble of the early eighteenth century. Prediction markets may have difficulty gaining broad organizational use if they are understood to work only ‘some of the time’ (consider the thought experiment of an engineering software package which works correctly ‘most’ of the time).

1.1.7 The Role of Feedback in Markets

Markets integrate varying levels of positive and negative feedback. If a market prices an asset below its intrinsic value then some market participants will buy that item causing the price to increase. Buying will continue until the price has risen close to the intrinsic value. This self correcting, negative feedback, loop is fundamental to the correct functioning of a market. However, the buying process, creating the rising prices, may trigger other participants to also buy; this additional buying creates even more price increases and can cause further buying. This represents positive feedback, which can continue for extended periods.

1.1.8 Motivations of the Research

The motivation of the present study is to better understand the underlying mechanisms and some of the design parameters of prediction markets in order to help organizations design more effective information tools. In particular, the research seeks to better understand the role of feedback in these markets.

1.2 Research Questions

The research questions focus on key aspects of prediction market use in organizational settings and are summarized below.

1.2.1 What are the Mechanisms Involved in the Information Assimilation Process?

What are the mechanisms by which markets assimilate information? What is the role of feedback in these processes? What types of failure mechanisms are possible and how should a prediction market organizer deal with them?

1.2.2 How does the Presence and Type of Feedback impact Market Performance?

As mentioned there are multiple aggregation methods. The incentivized poll worked extremely well in the Ox weight contest. A traditional market structure worked in the Challenger example. How do the factors of feedback structure and incentives impact performance? How should an organization choose?

1.2.3 How do Individuals use Information in the Allocation Process?

How are participants using private and public information in their decision processes and how does that impact the market performance? Can these strategies be modeled?

Table 1 summarizes the research questions and their relevance to the organizational prediction market designer.

Research Question	Relevance
What are the mechanisms involved in the information assimilation process?	How does a market achieve its results and how can it be managed? What are the error mechanisms in markets? And how can they be managed?
How does the presence and type of feedback impact market performance?	What is the relationship between prediction market designs and performance? Which structure should an organizational decision maker choose for their particular task?
How do individuals use information? What are the models for individual behavior?	Can the performance of the market be improved by changing participants' behavior? Who should be invited to participate?

Table 1. Summary of research questions.

2 Literature Review, Synthesis and Experimental Hypotheses

In this section the relevant literature is reviewed and synthesized to bring out the detail relevant to this work.

2.1 Market Background

This section reviews the important concepts from the study of markets.

2.1.1 Market Types

Many market types are available (cf. Harris, 2003 for more details). The exact dynamics will depend on the type of market as well as a host of other characteristics associated with that market - e.g., liquidity, presence of market makers, costs, etc. The discussion here will focus on three common methods: double auction; parimutuel; and fixed winnings.

2.1.1.1 Double Auction

Double auction markets are the most common form of markets in the financial world. They are also quite common in prediction markets – with the Iowa Electronic market being one example. Potential buyers and sellers submit bid and ask prices. When an overlap occurs (that is, bids greater than asking prices) a transaction can occur. The efficiency of a double auction market is dependent on the number of participants and the liquidity (frequency of buying and selling). For large capitalization companies in a major stock market (e.g., NASDAQ), double auction markets work extremely well. The spreads between, bid and ask prices are small, and many shares are traded so that the

prices are stable. However, for smaller markets, which typify many organizational prediction markets, double auction markets may be problematic. Double auction market processes have been studied extensively by traditional game theorists (cf. Gibbons, 1992) as well as complexity scientists (cf. Friedman & Rust, 1993).

2.1.1.2 Parimutuel Markets

Parimutuel betting systems have been developed as an alternative market structure for prediction markets. The parimutuel system as a prediction market was initially explored by Plott et al. (Plott, Wit, & Yang, 2003). Parimutuel systems are common in many gambling environments such as horse racing. In this system all bets are placed into a pool. The organizer removes a share (often called the take or vig) and the remainder is split amongst all winning tickets. The organizer often shows a running summary of the collected bets; for example, in the case of a horse race, the odds are calculated and displayed as the bets are received.

2.1.1.3 Fixed Winnings

Another market type, called here fixed winnings, would constitute a lottery which provides a fixed payout for each correct answer (Shelley, 1989). This type of market will be used in the present research.

2.1.2 Information Assimilation Capability of Markets

The concept of a market's ability to aggregate information goes back to some of the fundamental concepts of classical economics. Smith's (1776) invisible hand and Hayek's (1945) giant processing machine for decentralized information allude to the information processing capability of markets. However, the ideas were more fully

developed with the *Theory of Rational Expectations* and the *Efficient Market Hypothesis*.

Before reviewing these theories let's examine what economists mean by aggregation.

There are two tasks for the market:

- Bring together diverse sets of information and methods of interpreting information
- Make a judgment about what that information means

Some economists use the combined terms of *disseminate and aggregate* to convey the multiple tasks. However, in other cases, only aggregation is used. One could argue that the word aggregate may not precisely convey both functions so this work uses the term assimilate.

The theory of rational expectations states that agents, acting with complete access to the relevant information, forecast events in the future without bias. Any errors then would be due to random events. The theory is generally attributed to the American economist John Muth (1961). Grossman (Grossman, 1981; Grossman & Stiglitz, 1980) performed the theoretical analysis to show that under conditions of asymmetric information, the theory of rational expectations leads to effective aggregation of information by a market. The theory grew from a response to the theory of adaptive expectations. The *Theory of Adaptive Expectations* argues that future values are based on past values. However, the value cannot reach equilibrium under this assumption. So, rational expectations theory was developed to be consistent with equilibrium conditions. It is worth noting that the rational expectations hypothesis does not assume every decision of each individual is rational. Rather it assumes that the sum of all decisions by all market participants has no systematic bias and uses all relevant information in the formation of the expectation.

The theory of rational expectations can then be used to build the Efficient Market Hypothesis (Fama, 1965) . The efficient market hypothesis asserts that markets are efficient in terms of the prices, that is, the price reflects all known relevant information. If the theory is strictly correct, then it would not be possible to outperform the market. Since all known information is integrated into prices, only unknowable new information can affect prices and thus prices should be described by a random walk (a hypothesis made by Bachelier (1900) in his dissertation sixty years prior).

Plott and Sunder (1988) empirically showed that properly designed markets can indeed aggregate information. In their experiments, they defined three possible states of nature (X, Y and Z). If the correct state of nature (drawn randomly) was Z, then half the traders were told 'X is not correct' and the other half were told 'Y is not correct.' Thus, the collective information of the group was sufficient to precisely define the true state. They created double auction experimental markets for this research and showed that some of these markets disseminated and aggregated the information by correctly forecasting the correct state.

2.1.2.1 Theoretical Challenges to Classical Models

Several theoretical challenges to the classical models of market efficiency have been proposed. These are outlined below.

2.1.2.1.1 Keynesian Beauty Contest

Keynes (1936) questioned whether market participants were pricing based on fundamental values or beliefs about other market participants. He used the analogy of a

beauty contest to illustrate the point. In his fictional example he supposed a newspaper contest asked participants to predict the faces that would be judged the most beautiful to win a prize. A simple strategy is to choose faces that the entrant considered beautiful. However, a more sophisticated strategy would be to consider public perception and to attempt to predict the public's choice. An even more sophisticated strategy would be to assume other participants were also using this strategy and therefore the task is to predict what other participants think the public's perception will be. This line of thinking can be carried out indefinitely.

The Keynesian beauty contest represents another branch of game theory concerned with common knowledge (Binmore, 1991). A simple game sometimes called the number guessing game (or k game) was developed to test the ideas of the beauty contest. Each participant is asked to choose a number between 0 and 100. A prize is awarded to the participant whose number is closest to $2/3$ of the average. The first order reasoning might be an expectation that the average between 0 and 100 will be about 50 and $2/3$ of 50 is 33. However, if everyone were to pick 33, then the correct choice is 22. However, if everyone were to choose 22, the correct choice would be 15. The Nash equilibrium for this game is 0, and that's what a *rational* player should choose under the assumptions of common knowledge. Camerer, Ho and Chong (2004) reviewed 24 published experiments on the k -game. The lowest winning number in all the experiments was 15. There was some variation across class of participant, incentives and repetition but most participants only think a few levels deep.

2.1.2.1.2 The Efficient Market Paradox

The efficient market paradox, or information paradox, was originally proposed by Grossman and Stiglitz (1980). The paradox argues that if the markets reflect all information then there are no opportunities and no one will trade. However, markets require trading to be efficient since trading is the process which aggregates the information. So, if everyone believes the efficient market paradox, then it will not hold, as there will be no trading.

Sornette (2003) makes an interesting observation about this paradox:

“the more active and efficient the market, the more intelligent and hard working the investors; as a consequence the more random is the sequence of price changes generated by such a market. The most efficient of all is one in which price changes are completely random and unpredictable. ... Information leads to randomness, while lack of information leads to regularities.”

One argument is that the paradox can be solved from a dynamic perspective (Birchler & Butler, 2007) as a small time advantage can be sufficient incentive to acquire and disseminate new information (trade).

2.1.2.2 Practical Challenges to Classical Models: Bubbles/Mirages

Bubbles represent a significant challenge to classical models of market efficiency. The market work on bubbles and mirages will be reviewed.

2.1.2.2.1 Classical Work on Bubbles, Crashes and Mirages

Brunnermeir (2001) defines a bubble as a persistent deviation of an asset price from its fundamental value. This definition creates a conundrum with identifying a bubble in process since for so many assets there is no way to strictly identify the intrinsic value. Given that, many bubbles are only identified after a sudden drop in price (called a crash).

The logic behind this being that there was no sudden change in the asset or information associated with that asset which could have accounted for the sudden change, therefore the prior value was in fact a bubble.

Some argue that the existence of bubbles and crashes disprove the efficient market hypothesis. However, the efficient market hypothesis only asserts that all available information held by traders is aggregated in prices. There is no assertion by efficient market hypothesis that the information is fundamentally correct.

Charles Mackay (1932) spent three chapters discussing economic bubbles including the South Sea Company and Mississippi Company bubbles of the early eighteenth century and the Dutch tulip mania of the early seventeenth century. Burton Malkiel's *A Random Walk Down Wall Street* (2003) makes spirited references to Mackay's descriptions. As a more recent example, the dot-com stock rise and fall of the late 1990's is described as a bubble (Shiller, 2000).

2.1.2.2.2 Bubbles in Experimental Economics

Bubbles have been repeatedly created in the markets of experimental economics (cf. Porter & Smith, 1995; V. L. Smith, Suchanek, & Williams, 1988). In these experiments the conundrum over fundamental values can be explicitly addressed. For example, some of the securities used in the trading experiments had precise values through their experimental lifetime. Miller (2002) provides a synthesis of these experiments and observes that momentum and excess capital as two common contributing factors to bubbles. For example, the movement of price towards its fundamental value provides

the momentum which can create an overshoot bubble. He also observes that increased availability of trading capital (fiat money given to participants) leads to both the likelihood and size of bubbles.

2.1.3 Prediction Markets

Recent books (Abramowicz, 2007; Hahn & Tetlock, 2006) and an extensive literature review (Tziralis & Tatsiopoulos, 2007) are available on prediction markets. Only a small sampling of this literature will be discussed here.

Tziralis and Tatsiopoulos (2007) reviewed 155 articles published between 1990 and 2006. These span a wide range of topics including theoretical work, market description, applications and legal or policy concerns. They found the pace of publication to be increasing with very high growth in the last 5 years.

The Iowa Electronic Market (IEM), begun in 1988, is the most recognized and earliest example of a prediction market (R. Forsyth, Nelson, Neumann, & Wright, 1992). Established by the business school at the University of Iowa, this is a futures market for political events – with greatest notoriety around US Presidential elections. As mentioned the markets have been shown to be very accurate and consistently predict better than more traditional methods such as polling (J. Berg et al., 2008).

Incentives are fundamental in economics. Many organizational prediction markets offer a monetary incentive (Hopman, 2007). Other organizational markets offer (Cowgill,

Wolfers, & Zitzewitz, 2008) prizes (such as company logo shirts or caps). Both have been shown to be effective. It is worth noting that employees participating in a company prediction market have indirect incentives. For example, if the results of the market are used by the management and this helps the project to succeed then the company may improve its overall results. This could benefit participants in various ways ranging from securing their own job to adding to their financial interests (e.g., in terms of bonuses, stock options, etc.).

2.1.4 Experimental Prediction Markets

Of prime interest in this area are the experimental prediction markets organized and reported by Charles Plott and colleagues at The California Institute of Technology (Caltech). These works followed the framework of experimental economics in they attempt to control the information given to participants. In these markets, participants, almost exclusively Caltech students, are given controlled information which is primarily balls drawn from an urn.

A brief description of the urn markets will be given. Participants are given private data and can observe how other participants are acting (public data). The private data consists of a set of draws from an urn. In what they describe as the Probabilistic Information Condition (PIC) the urn contains 15 balls. The balls consist of 6 types representing the states of nature – here considered simply to be letters A through F. Prior to the prediction market (PM) the researcher randomly chooses one of the balls to be the correct state. This selection is not revealed to the participants. Suppose the

random selection chooses B to be the correct state. The urn would then be populated with 5 B balls, and 2 balls of each of the other types. Participants know the relative percentages but do not know which state is correct. Each PM participant is then given the results of three draws from the urn, with replacement. The result of these three draws, the draw set, is the private information of each participant. The participants are given an allocation of fiat which can be used to purchase tickets across the six states. The cumulative purchased tickets are displayed and available to the participants – the public information. After the PM closes, and the correct state is revealed, the pot is split evenly across the tickets purchased for the correct state. Any unused fiat is returned to the researcher (so there is no incentive to hold fiat).

Plott, Wit and Yang's paper "Parimutuel betting markets as information aggregation devices: experimental results" remains the most important work on experimental parimutuel markets (2003). Plott and team ran two types of experiments. In the first, balls were drawn from an urn and results given to participants as described above (PIC). The second type of experiment was called 'Not sets.' Here, there were again six states of nature, but participants were given definitive information about what state was not correct. Enough information was supplied such that the group, as a whole, could precisely determine the correct state if they were able to share all their data. Participants were given an allocation of fiat money with which to purchase tickets. Once purchased, tickets could not be refunded or exchanged¹. Once the market was complete, the winning pot was split evenly across all correct tickets. The numbers of participants in

¹ This eliminates the possibility of a crash

the markets ranged from 10-15. The markets were open for between 4 and 5 minutes. Participants were told that they could randomly close any time after 4 minutes. Plott and team conclude that information aggregation can be detected in both types of markets. They also observed mirages (defined as the mode of the distribution being on an incorrect state), which occurred 9 times in the 38 markets they ran (23.4%). The team observed that the majority of the tickets were allocated in a small time window just prior to the earliest time the markets could close. They felt this behavior inhibited the aggregation of information. Further, they observed bluffing and long shot bias in the markets, although they noted these effects were small.

In a second set of experiments at Caltech, Axelrod et al. (Axelrod, Kulick, Plott, & Roust, 2007) made two modifications to the market structure: 1) a cost for waiting was implemented by increasing the price of tickets with time; 2) opening a second market immediately after the first market had finished. The team theorized that some of the observed effects such as the long shot bias were the result of disequilibrium which arises from the fact that tickets cannot be resold. So, the second market was opened after final odds from the first market were published to the group, but before the correct state of nature was identified. They found that both of these mechanisms improved the ability of the market to aggregate information as compared to the original mechanism. However, mirages continued to occur at similar rates².

² In the non-repeated market mirages occurred in 29/90 markets (32%) in the repeated market mirages occurred in 8/42 markets (19%)

In another set of experiments at Caltech, Roust and Plott (2005) designed a two stage market. First, participants were given private information as before. They were then given an allocation of fiat to purchase tickets in a market without feedback (that is, it was an incentivized poll). The results of the first phase were then published to the group without identification of the correct state. Then, a second market, with feedback, was conducted as described in the Axelrod experiment, with an increasing cost of tickets as the experiments proceeded. Table 2 shows the results. Overall, they observed approximately the same rate of mirages (27/90 or 30%). It is interesting to notice that mirages were seen in the first round (incentivized poll). However, in 70 of the 90 markets they only had 10 participants. This small number could lead to statistical anomalies in the ball draw

First Stage Prediction	Second Stage Prediction	Count
Correct	Correct	63
Incorrect	Incorrect	18 ³
Correct	Incorrect	1
Incorrect	Correct	4

Table 2. Summary of results from Roust & Plott Experiments (2005)

³ Of the 18 where both first and second stage were incorrect, 14 predicted the same state while 4 predicted different states

2.2 Individual Decision Processes

Aspects of the subject's decision processes are examined in this section. The focus will be on the processes involved in experimental prediction markets.

2.2.1 Overview

Decision frameworks and the overall steps in the allocation process are discussed. Figure 1 shows an overview of the steps in the decision process in the experimental prediction market used in this study. The three stages include: gathering of data including the private draw and public information if available; assessment of the probability that any given state is the correct state; and, finally a decision of how to allocate tickets.

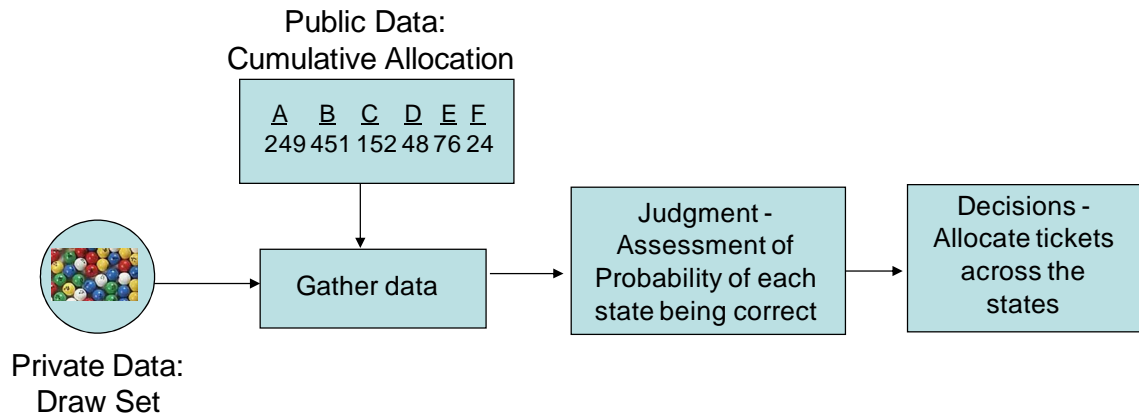


Figure 1. Steps in the ticket allocation process

There are a number of excellent sources of review of existing work in judgment and decision making. Hastie and Dawes (2001) provides an excellent overview of key issues in judgment and decision making. Resnick (1987) also provides an excellent overview of decision theory. Camerer (1995) provides an excellent overview and specifically highlights areas relevant to economics. Baird (1989) gives an excellent

overview of key decision analysis tools. Camerer's book *Behavioral Game Theory* (2003) gives an excellent review of research on descriptive behaviors in decisions under conflict.

2.2.2 Judgment - Assessment of Probability

After collecting their data, participants must make an assessment of the probability of each state being the correct state.

2.2.2.1 Bayes' Theorem

Using only the private information, Bayes' Theorem can be used to calculate the posterior probability that any state is the correct state given any combination of draws from the urn. Bayes' Theorem is shown below -

$$p(H_{post} | D) = \frac{p(H_{prior}) \times p(D | H)}{p(D_{independent_of_hypothesis})}$$

Where:

$p(H_{prior})$ = prior probability of the hypothesis

$p(D|H)$ = conditional probability of the data given the hypothesis

$p(D)$ = probability of the data = $p(H)p(D|H) + p(-H)p(D|-H)$

$p(H_{post}|D)$ = posterior probability of the hypothesis given the data

The probabilities for the Plott PIC market are calculated and shown in Table 3. There are three possible types of draws: all three balls match (e.g., AAA, BBB, CCC, etc.); Two balls match (e.g., AAB, AAD, BBC, etc.); or none of the balls match (e.g., ABE, CDF, AEF, etc.). These three types are shown as the rows in the table. For the first row, all balls match, then the posterior probability that the ball type received is the correct state is

calculated to be 75.8% (that is, if the experiment of drawing balls from the urn were repeated a large number of times, 75.8% of the time that three matching balls was received, the type of ball received would be the type with 5 balls in the urn – the correct state). Still, it is possible to draw three balls of the incorrect state (that is, with two balls in the urn). This will happen 4.8% of the time. Again, the urn contains 5 balls of the correct state of nature and 2 balls of each of the other states.

	Most Frequent (e.g., A)	2 nd Most Frequent (e.g., B)	2 nd Most Frequent (e.g., C)	Each Not Drawn (e.g., DEF)
Three matching draws (e.g., AAA)	75.8%	4.8%	4.8%	4.8%
Two matching draws (e.g., AAB)	49.0%	19.6%	7.8%	7.8%
Three different draws (e.g., ABC)	23.8%	23.8%	23.8%	9.5%

Table 3. Bayesian posterior probabilities of private information

To illustrate the calculation consider the situation where a participant receives a draw of 2

A balls and 1 C ball. What is the probability that A is the correct state?

$$p(H_{prior}) = 1/6 \text{ (all equally probable)}$$

$$p(D/H) = \text{probability A given A is the correct state (i.e., Urn is biased for A)} \\ = (1/3)^2 * (2/15)^1 * (2/15)^0 * (2/15)^0 * (2/15)^0 * (2/15)^0$$

$$p(D_{independent \text{ of hypothesis}}) = \\ (((1/6) * (1/3)^{\#A} * (2/15)^{(Total - \#A)}) + \\ ((1/6) * (1/3)^{\#B} * (2/15)^{(Total - \#B)}) + \\ ((1/6) * (1/3)^{\#C} * (2/15)^{(Total - \#C)}) + \\ ((1/6) * (1/3)^{\#D} * (2/15)^{(Total - \#D)}) + \\ ((1/6) * (1/3)^{\#E} * (2/15)^{(Total - \#E)}) + \\ ((1/6) * (1/3)^{\#F} * (2/15)^{(Total - \#F)}))$$

Where, Total = Total number of balls drawn = 3
and, #A=2, #C=1, All other = 0

$$\begin{aligned}
p(H_{prior}) &= 1/6 \\
p(A/H) &= .0148 \\
p(D_{independent\ of\ hypothesis}) &= .0025 + .0004 + .0010 + .0004 + .0004 = .0050 \\
p(A/D) &= ((1/6) * .0148) / .0050 = .4902
\end{aligned}$$

2.2.2.2 Departures from Bayes' Theorem

Several research studies have shown significant deviations from Bayes' theory in practice.

2.2.2.2.1 Base Rate Neglect

The phenomenon of base rate neglect was first outlined in a famous experiment by Kahneman and Tversky (1973) and repeated by Maya Bar-Hillel (1980). A story approach was used to describe the situation of an accident involving taxi cabs where participants overwhelmingly demonstrated a neglect for the base rate in the probability assessment responses.

El-Gamal and Grether (1995) performed an extensive bingo cage experiment to test the usage of Bayesian rules. Two bingo cages with different distributions of balls were randomly selected by the draw of a ball from a third cage. The distribution of balls was varied in the initial cage and the usage of this *prior* probability was studied. They found that the most prominent rule being used was Bayesian. However, they did see significant numbers of subjects which used alternative rules. The second most common rule they saw in the subjects they called *representativeness*. Here participants based their choices on how the ball draw matched the cage composition, and thus essentially neglecting the ratio of balls in the first cage.

2.2.2.2.2 Conservatism

Experiments by Phillips and Edwards (1966) saw that subjects behave conservatively (i.e., update posterior probabilities insufficiently on the basis of new information) under various experimental conditions. They found:

"... conservatism was unaffected by prior probabilities, remained constant as the amount of data increased, and decreased as the diagnostic value of each datum decreased."

In the El-Gamal and Grether (1995) experiments mentioned in section 2.2.2.2.1 in the third most prominent rule observed, which they called *conservatism*, subjects over weighted the data based on the composition of the first cage.

A large number of studies have shown a common bias being overestimation of the probability of low frequency events and underestimation of the probability of high frequency events (Fischhoff, Lichtenstein, Slovic, Derby, & Keeney, 1981; Hastie & Dawes, 2001; Viscusi, 1985). For example, Hastie and Dawes (2001) show a graph of the statistical estimate of deaths per year for various causes vs. the judgmental estimates by participants. Very low frequency events are shown to be overestimated by several orders of magnitude. At the other end of the scale, subjects were seen to underestimate common causes of death by several orders of magnitude.

Anchoring and adjustment is a potential mechanism for the conservative behavior (Tversky & Kahneman, 1974). Subjects can anchor on any initial information and fail to adjust for the task.

2.2.2.2.3 Pseudodiagnosticity

Pseudodiagnosticity is a consequence of the failure to consider alternative hypotheses in a probabilistic inference task (Dougherty, Mynatt, Tweney, & Schaivo, 1979). It is evaluating the diagnosticity of information on the basis of only one of the required conditional probabilities. Eddy (1988) found 95% of practicing physicians grossly miscalculated the inferred incidence of cancer in his experiment. Eddy used the following statement in his experiment:

“The prevalence of breast cancer is 1% for women over age 40. A widely used test, mammography, gives a positive result in 10% of women without breast cancer, and in 80% of women with breast cancer. What is the probability that a woman in this age class who tests positive actually has breast cancer?”

While the correct answer is 7.5% as calculated using Bayes' theorem, 95 out of 100 physicians responded “about 75%.” Eddy attributed this to a failure to consider the alternative hypothesis.

2.2.2.2.4 Other Factors

Gigerenzer and Hoffrage (1995) have studied the effect of data shown as frequency versus percentages in Bayesian probability tasks. Their research shows more nearly Bayesian behavior with frequencies than with percentages or probabilities.

2.2.2.3 Integrating Multiple Data Sources

Information integration theory was developed by Norman Anderson (1981) and seeks to describe how information is integrated in the judgment and decision process. The theory asserts individuals go through a valuation phase followed by an integration phase. The valuation focuses on the psychological judgment of value or probability as opposed to the corresponding observable measures. Integration involves the method the decision maker uses to combine multiple inputs. The theory argues that individuals use simple algebraic rules in the integration processes.

Anderson describes a relevant experiment with two urns. Each has a specified proportion of red and white balls. One of the two urns will be selected with a given probability. Subjects were asked the probability that a white ball would be drawn given a mix of red and white balls in the two urns. The probability model is:

$$\text{Prob(White)} = \text{Prob(Urn A)}\text{Prob(White|Urn A)} + [1 - \text{Prob(Urn A)}] \text{Prob(White|Urn B)}$$

Anderson's results showed basic agreement between the subjects and the predicted probability. He suggested that this supported his cognitive algebra concepts where the subjects are able to effectively perform multiplication and addition.

2.2.3 Decisions - Allocation of Resources

Once the participant has assessed the probability that any state is the correct state they must make a decision on how to allocate their tickets. In this section relevant theories are discussed.

2.2.3.1 Expected Utility Theory

Classical theories of choice under risk focus on the *Expected Utility Theory*. Bernoulli (1738) originally described the theory but it gained prominence when von Neumann and Morgenstern (1944) were able to derive it from basic axioms. Simply stated, the theory argues that choice between probabilistic options can be made by comparing the weighted sums of the individual utility values multiplied by their respective probabilities. The theory supports choices where there are multiple selections down any branch of the decision tree, by summing up the products of utility and probability of all choices down a branch. In many classical theories the concept of rational decisions was tied to the maximization of expected utility (i.e., choosing the option with the highest expected utility). von Neumann and Morgenstern's axioms and expected utility theory have been extremely valuable in the development of decision theory and game theory by making it possible to formulate certain problems mathematically.

2.2.3.2 Departures from Expected Utility Theory

Several of the most relevant departures from expected utility theory will be discussed. In addition, some of the theories which have been proposed to account for this behavior will be described.

2.2.3.2.1 Dependence on Irrelevant Alternatives

A number of paradoxes have been used to test and question the expected utility theory.

One of the most well known is the *Allais Paradox* (1953) as illustrated below in Table 4.

The studies showed that subjects may not adhere to the independence axiom.

Experiment #1

	Probability	Payout
Choice 1	100%	1M Francs
Choice 2	89%	1M Francs
	1%	0
	10%	5M Francs

Experiment #2

	Probability	Payout
Choice 1	89%	0
	11%	1M Francs
Choice 2	90%	0
	10%	5M Francs

Table 4. Illustration of choices from Allais paradox

In Experiment 1 participants tend to choose Choice 1. In Experiment 2 participants tend to select Choice 2. The Allais Paradox challenges the expected utility theory by showing the participants do not strictly adhere to the underlying axioms. By selecting choice 1 in experiment 1 and choice 2 in experiment 2 the independence axiom is violated (cf. Hastie & Dawes, 2001 for the derivation).

2.2.3.2.2 Ambiguity Aversion

In the Ellsberg paradox (1961) participants must choose between two gambles with the same expected value. In one gamble there is a known distribution and with the other gamble there is an unknown distribution. Participants overwhelmingly choose the gamble with the known distribution. This displays an aversion to ambiguity (or uncertainty).

2.2.3.2.3 Preference Reversals

The phenomenon of preference reversals is based in the work of Slovic and Lichtenstein (1968) with extensive review and discussion in the recent edited volume (Slovic & Lichtenstein, 2006). In a preference reversal two lotteries are constructed wherein many participants place a higher price on one lottery while preferring to participate in the other lottery. For example, lottery 1, which is typically referred to as a P bet (P standing for probability - these bets have a higher probability of winning), might offer a 80% probability to win \$2 and 20% probability to lose \$1. An example of lottery 2, which is typically referred to as a \$ bet (\$ bets having a higher winning dollar amount), might be 20% chance to win \$9 and 80% chance to lose \$0.50. When participants are asked which lottery they would like to participate in, they usually prefer P bet to \$ bet lotteries. However, when asked how these lotteries should be priced, participants feel that \$ bet lotteries should be priced higher than P bet lotteries. In one study Slovic and Lichtenstein found that 73 percent of participants made the reversal. Concerned about the implications of this phenomenon to economic theory Grether and Plott (1979) systematically tested a number of methodological consideration and concluded the results were robust. In addition to challenging expected utility theory as a descriptive model, the preference reversal phenomenon illustrates an important framing issue.

2.2.3.2.4 Satisficing

In the theory of Satisficing developed by Simon (1976), the cost of maximizing is taken into consideration. For example, an individual may choose a satisfactory, rather than optimal solution, if their search time is limited.

2.2.3.2.5 Prospect Theory

Prospect Theory, by Kahneman and Tversky (1979), develops a framework around expected utility and models a number of departures from expected utility theory with numerous empirical studies supporting it. The descriptive theory describes how individuals evaluate potential loss and gains in risky decisions. Hastie and Dawes (2001) describe three key components of the utility function assumed by the theory:

1. *"An individual views monetary consequences in terms of changes from a reference level - which is usually the individual's status quo."*
2. *"The resulting value function is steeper for losses than for gains - implies loss aversion."*
3. *"The curve is concave for gains and convex for losses, implying the decision maker will be risk averse when choosing between gains and risk seeking when choosing between losses."*

In addition, Prospect Theory (a) substitutes decision weights for probabilities, with decision weight being a non-linear function of probability, and (b) assumes a preliminary editing phase in which a reference point is set and outcomes may be combined or segregated.

2.3 Interactions in Markets

Markets by their very definition involve multiple agents. This section reviews relevant interaction effects from the literature.

2.3.1 Rational Markets with Irrational Agents

Given the observed irrational behavior by decision psychologists and behavioral economists, how do classical economists reconcile their views around rational expectations and efficient markets? Camerer (1992) lists four hypotheses.

First, the *cancellation* hypothesis would assert that individual errors are random and cancel each other. However the cancellation hypothesis requires the judgments be randomly distributed around the rational judgment while the behavioral research shows systematic errors.

Second, the *smart few* hypothesis asserts the market actions are dominated by a few active traders who are generally rational. This theory requires some level of self insight into the error mechanisms by the participants.

Third, the *learning* hypothesis asserts traders learn from their experiences or buy advice which averts the irrational behavior. Counterarguments include evidence that learning from individual experience is difficult (Einhorn & Hogarth, 1978) and imitation learning can be even more difficult.

Finally, the *evolutionary* hypothesis argues that traders who make rationality errors are selected out – that is, they go bankrupt because of their errors. Thus, only unbiased traders would remain. The arguments against this theory are that economic evolution

could work very slowly. Second, evolutionary processes may not have any inherent bias for or against rational decision processes.

2.3.2 Feedback

Feedback is the engine of market interactions. In this section, various ideas discussed in the literature are reviewed.

2.3.2.1 Information Cascades

Information Cascades have been proposed as an explanation of herding behavior in markets (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992). To understand this mechanism consider a simple example. Individuals must choose between two doors, A and B, one correct and one not correct. They each are given private probabilistic information on which is the correct choice. This private information is not shared directly however their decisions are visible to the other participants. Consider the situation where the private information has probability of 2/3 of being correct. You are the third person to make a choice and your information indicates B has a higher probability of being the correct choice. However, you see the first and second person choose A before you must make your selection. Assuming everyone is using Bayes' theorem one might reason as follows: the prior probabilities are 50/50 and the first person in line had no information other than their own private information so application of Bayes' theorem would suggest they follow their private signal. So, that would imply the first person had private signal which indicated door A was correct. The second person would have also assumed that the first person's private signal was door A. If their private information was door B Bayes' theorem would suggest that both doors have equal

probability. Making a second assumption that in the case of equal posterior probabilities the second person makes their prediction based on their private signal indicates that by also choosing door A, their private signal was also door A. Now it is your turn to choose. If the two previous private signals were indeed door A, then application of Bayes' theorem would suggest ignoring your private signal of door B and choosing door A. At this point the information cascade has begun. All the following individuals have the same data from the first two participants as well as all who preceded them who would have continued to choose door A. Information cascades have been extensively studied and have a rich literature (cf. Hirshleifer & Teoh, 2003 for a review).

Experiments have validated (L. R. Anderson & Holt, 1997), in general, the predictions of information cascade theory. There are several important caveats to these experiments however. First, they represent a carefully controlled experiment where the known probabilities are available to all participants (in this case, balls are selected from urns with known frequencies). So while the participants did not know the exact information of others, they did know the nature of that information. This differs from the 'real world' in that market participants have no idea of the nature of others' information. The second caveat is there are always a few participants who do not act in accordance with the theory (e.g., do not take the Bayesian rational option). We know that some individuals make all manner of deviation from rationality in decision processes – and this result simply affirms.

The information cascade represents a classic example of a Bayesian game (Harsanyi, 1967, 1968). However, this perspective is rarely mentioned in the literature. Eyster and Rabin (2009) discuss this aspect in their work, however they do not present any prescribed strategies.

2.3.2.2 Reputational Herding

Scharfstein and Stein (1990) propose a model for herd behavior based on managers wishing to support their reputation as decision makers and mimicking the investments of other managers in order to minimize the risk of uniquely making a bad decision. This theory relies on some visibility into the manager's holdings (e.g., a fund manager who periodically publishes their investments to managers or shareholders). Graham (1999) suggests this mechanism in herding behavior seen amongst investment newsletters.

2.3.2.3 Crowd Behavior

Plummer (2003) has suggested that crowd behavior is important in the actions of markets. Forsyth (1999) provides an excellent overview of crowd or collective behavior theories. He suggests a definition of collective (or crowd) behavior as the actions of a large group of people who are responding in a similar way to an event or situation (based on McPhail, 1991). Some theorists believe crowds have unique properties which differ from the dynamics present in smaller groups. Forsyth discusses four key theories on crowd behavior: contagion theory; emergent norm theory; convergence theory; and, deindividualization theory

Le Bon (1895) originally proposed *contagion theory* where emotions and behaviors were transmitted from one crowd member to others. He proposed that this process could

transform the crowd from thoughtful and rational individuals to the extreme behaviors sometimes observed.

Turner and Killian (1972) proposed *emergent norm theory* which argues that the collective behavior is caused by adherence to group norms created spontaneously by the crowd.

Convergence theory asserts that the crowd's collective behavior is present, either direct or latent, in the group beforehand.

Deindividuation theory (Festinger, Pepitone, & Newcomb, 1952; Zimbardo, 1969) asserts that crowd factors such as anonymity, arousal, etc. create a lowered threshold of normally restrained behavior.

Forsyth notes the behavior need not involve groups that are physically together as psychologists list fads, fashions, rumors, mass hysteria and social movements as exhibiting crowd behavior. Notably, markets would fall under this category.

Many recent studies have concluded that dramatic changes in self-control and atypical behavior are actually rare. Forsyth (1999) writes:

“They concluded that the conditions that exist in mobs, such as membership in a large crowd, anonymity, and a decline in self-awareness, prompt group members to conform to situation-specific group norms, but that these norms are rarely inconsistent with general social norms (Postmes & Spears, 1998).”

2.3.2.4 Feedback in Delphi

The *Delphi method* (Linstone & Turoff, 1975) is a forecasting tool which uses feedback.

With the Delphi method a panel of experts is asked to provide comments which a facilitator anonymously combines and republishes to the panel. The experts are asked to review the comments and forecasts from the other experts, update their thinking if appropriate, and resubmit their comments. The process continues until the results are stable or a set number of rounds have occurred.

Of particular interest is the impact of feedback. Are the groups in effect converging to the correct answer, or just consensus? An excellent meta-analysis is done by Rowe and Wright (1999). They summarized the results across 16 papers. There were 21 experiments in these papers and they saw divergences (where the group moved away from the correct answer) in 3 cases.

2.3.2.5 Positive and Negative Feedback in Markets

A feedback loop is a series of connections within a system which affects itself (Sterman, 2000). The type of feedback, either positive or negative, is determined by the type of the effect. In a system with a positive feedback loop, a perturbation at the input of the system will be increased as it travels around the loop. In a system with negative feedback the perturbation will be decreased as it travels around the loop.

In the perspective of the efficient market hypothesis the price provides negative feedback to the market. As market participants assess their information, they develop an expectation of the price. If the market price they observe is greater than this expectation

they will take actions which tend to decrease the market price (that is, they may sell the item if they own it). By the same reasoning, if their expectation is lower than the current market price, they will take actions which tend to increase the market price (that is, they may buy the item). By contrast, a system where an information mirage is underway is in a positive feedback cycle. As more participants make the same choice it encourages subsequent participant to make the same choice.

2.3.3 Systems Theories of markets

A number of theories and analyses of markets have been developed within the systems science field. These will be briefly described.

2.3.3.1 System Dynamic Market Models

System dynamics can be used to model markets. Sterman (2000) discusses system dynamic models of commodity cycles. These models encompass supply and demand factors and endogenously create price cycles. One of the important lessons of system dynamics is that very simple structures can create complex dynamics. For example, the numerous varieties of S type growth (e.g., exponential growth, limiting effects, oscillations, etc.) all come from a structure with two simple feedback loops.

2.3.3.2 Catastrophe Theory

Zeeman (1974) proposes an alternative model for bubbles and crashes associated with catastrophe theory. Zeeman considers the market to be a gradient system and argues the dynamics of the interaction between fundamentalists and chartists (technical traders) gives rise to crashes that can be modeled with a cusp catastrophe. These concepts are also discussed in depth by Rosser (1991).

2.3.3.3 Agent Based Models

Various studies have created agent based market models. Cohen et al. (1983) provided one of the earliest reports. LeBaron (2000) provides a review of various work. The most extensive effort was carried out by the Santa Fe institute (W. B. Arthur, Holland, LeBaron, Palmer, & Tayler, 1997).

2.3.3.4 Markets as Complex Evolving System

A theory of markets as complex evolving and adaptive systems has been proposed (cf. Beinhocker, 2006), which has been predominately studied at the Santa Fe Institute (P. W. Anderson, Arrow, & Pines, 1988; W. Brian Arthur, Durlauf, & Lane, 1997). While not precisely defined, this perspective incorporates many of the traditional systems science concepts:

- Nonlinear dynamics – As opposed to traditional economics, equilibrium is not assumed. In fact, dynamics are inherently incorporated into the theories. Furthermore, linearity is not assumed.
- Interactions and emergence – macro phenomenon emerge from the interactions of agents within the system
- Evolutionary – Multiple aspects (behaviors, strategies, actions) of agents evolve in response to the environment.

Closely related is the developing field of Econophysics (c.f. Johnson, Jefferies, & Hui, 2003; Mantegna & Stanley, 2000). In the Econophysics framework, none of the typical assumptions made in classical Economics are used (e.g., normality of distributions, random walk behavior, rationality of agents, etc.). Instead, Johnson et al. (2003) describe the basic tenants as: feedback; non-stationarity; many interacting agents; adaptation; evolution; single realization; and, open systems.

2.4 Decisions Frameworks in Economics

Two key frameworks are relevant to the present study – behavioral economics and experimental economics.

Behavioral economics (or the closely related field of behavioral finance) is the integration of economics with the building blocks from psychology and decision science.

Assumptions around the aggregate rationality of economic agents are suspended and the impacts of behaviors observed in human decision processes on economic systems are studied. The initial work in the field was strongly influenced by the development of *Prospect Theory* by Kahneman and Tversdsky (1979) and *Bounded Rationality* by Simon (1976). An excellent overview is given in *Inefficient Markets: An introduction to Behavioral Finance* by Shleifer (2000).

Experimental economics relies on the application of standard experimental methods to study economic issues. Thus, experimental economics also takes a decidedly descriptive view of human behavior. Camerer (1995) provides an excellent review with illustration of the viewpoint of experimental economics.

One important difference between behavioral and experimental economics is in their inquiring systems. Behavioral economics primarily relies on analysis of the vast data sets available from actual economic systems in its analysis. This Lockean form of inquiry (Mitroff & Turoff, 1973) is in contrast to the Kantian form used in

experimental economics which relies on experimental markets where variables of interest are precisely controlled for focused studies of cause-effect relationships.

3 Experimental Hypotheses and Methods

The experimental hypotheses and methods are discussed.

3.1 *Synthesis and Experimental Hypotheses*

The literature is synthesized and a set of experimental hypotheses are developed to address the research questions set forth in section 1.2.

3.1.1 Research Framework and Market Measurements

The focus of the research and some of the key tools will be described.

3.1.1.1 Research Framework

The framework of experimental economics will be adopted for the present study. All information to participants will be known by the researcher; and all participants' actions will be captured.

An experimental prediction market similar to the Plott et al (2003) PIC market will be used. Participants will be able to allocate a fixed number of tickets across these six states. The cumulative allocation of tickets per state will be the market result.

3.1.1.2 Entropy

The entropy of the final distribution of allocated tickets can be used as a measure of the certainty, or the conviction, of the prediction. The entropy is computed as shown in the equation below:

$$H(x) = - \sum_{all\ x} (p(x) * \log p(x))$$

where $p(x)$ is the probability associated with a particular state and is simply the tickets allocated to the state divided by the total number of allocated tickets. The log nature of Entropy tends to compress differences. The minimum value of entropy is 0 which occurs when all tickets are allocated on one state. The maximum entropy is 2.585 which occurs when tickets are evenly distributed across the six possible states.

3.1.1.3 Evaluating two Distributions - Wurtz Distance

Several descriptive tests have also been used to describe differences in distributions (Plott et al., 2003). The Wurtz number will be used in the present study (Wurtz, 1997) as defined by:

$$W(p, q) = 0.5 \sum_{i=1}^K |p_i - q_i|$$

where p and q are the probabilities of each state (that is, the observed ticket allocation divided by the total). This metric was chosen because it is well behaved for allocations of 0 tickets.

3.1.1.4 Aggregate Information Available (AIA)

Plot et al. (2003) describe a measure they termed Aggregate Information Available (AIA).

AIA is simply the Bayesian posterior probability if all draws which have been distributed are known. In the market used for the present research, the AIA is a distribution of probabilities across the six states.

Of particular interest is the Wurtz distance between a given cumulative allocation in a market and the AIA distribution. This gives a measure of how close the allocation is to perfect assimilation of information.

3.1.1.5 Statistical Significance of a Result

The cumulative allocation of tickets reflects the market's results. A two stage statistical test for significance of the results will be employed for the present study. In the first phase the accumulated results will be compared to a uniform distribution to determine if there is any information in the distribution. In the second phase the two states with the largest distributions will be compared for significance of the mode. This method is analogous to the use of ANOVA to test for significance on a group of means and then post-hoc tests to determine which means are statistically different. The method is illustrated in Table 5. In this example, the cumulative allocation across states is 10,20,100,40,80,50. With a total of 300 tickets allocated, this allocation is compared to a uniform allocation of 50 tickets per state. With standard categorical data analysis the resulting p-value obtained from the Chi-square distribution is less than 0.05, so the allocation is deemed to have information content. In the second phase, the mode of 100 tickets on state C is tested against the second highest allocation, 80 tickets on state E. These are compared against a uniform distribution of the tickets across the two state (in this case 90 tickets on each). Here categorical data analysis yields a p-value which is greater than 0.05. Therefore, in this example, the mode is not significantly different than the second highest allocation.

	A	B	C	D	E	F	Total	Chi-Square p-value
Actual	10	20	100	40	80	50	300	
Uniform	50	50	50	50	50	50		3.13858E-24
Actual			100		80			
Uniform			90		90			0.136037165

Table 5. Illustration of two stage evaluation of allocations

This process was used to describe a result as being in one of three possible states:

1. Not Significant - p-value of first or second test is $> .05$
2. Correct - p-value of both first and second test is $< .05$ and the mode is correct
3. Incorrect - p-value of both first and second test is $< .05$ and the mode is incorrect

3.1.1.6 Decisiveness and Correctness

In terms of a market's performance, two interesting metrics can be calculated using the categorization of Correct, Incorrect and Not Significant:

$$\text{Decisiveness} = \text{Number Significant} / \text{Total Number of markets}$$

or

$$\text{Decisiveness} = (\text{Number Correct} + \text{Number Incorrect}) / \text{Total Number of markets}$$

$$\text{Correctness} = \text{Number Correct} / \text{Number Significant}$$

or

$$\text{Correctness} = \text{Number Correct} / (\text{Number Correct} + \text{Number Incorrect})$$

Decisiveness is a representation of the market's ability to make a judgment. Correctness is the probability that this judgment is correct. Both are important considerations for a prediction market designer.

Decisiveness and correctness values from two different test cases will be compared for statistically significant differences using the Fisher's exact test on the contingency tables shown in Table 6 and Table 7

	Case 1	Case 2
Significant	Number Correct in case 1 + Number Incorrect in case 1	Number Correct in case 2 + Number Incorrect in case 2
Not Significant	Number Not Significant in case 1	Number Not Significant in case 2

Table 6. Contingency table to test differences in decisiveness .The subscripts refer to case number

	Case 1	Case 2
Correct	Number Correct in case 1	Number Correct in case 2
Not Correct	Number Incorrect in case 1	Number Incorrect in case 2

Table 7. Contingency table to test differences in correctness

3.1.1.7 Evaluating Prediction Markets - Accuracy and Precision

Typically, a measurement system is judged by two characteristics: precision and accuracy.

Precision is a measure of the repeatability or variance in the measurements. Accuracy is a measure of the degree to which the correct quantity is reported. As shown in Figure 2 these two need not be related. The best measurement system is both accurate and precise, the worst is neither.

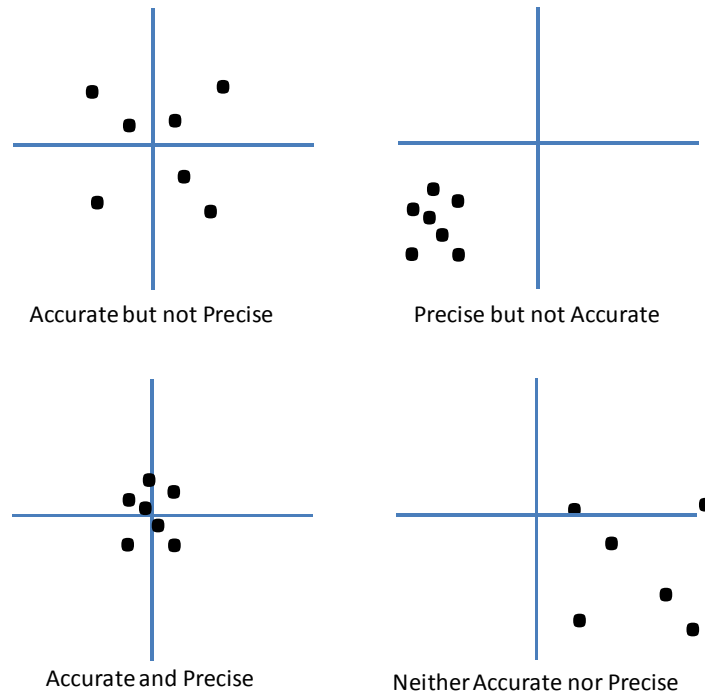


Figure 2. Illustration of accuracy vs. precision. Each dot represents a different measurement. The axis in the plots represent two variables and the correct value is the origin.

Table 8 summarizes the market measures discussed in this section and whether they characterize the precision or accuracy of the market. The correct, incorrect, not significant parameter characterizes both the precision and accuracy of the market. In the present study, since the researcher has full knowledge of all relevant information, these parameters can be computed as the market is progressing. In practice, a prediction market organizer can only make this characterization when the market is complete and if the state of the phenomenon the market was trying to predict is known. The Wurtz distance between the cumulative allocation and AIA distribution is another valuable measurement of both the market's precision and accuracy. In the present study, since the draw information is always known, the AIA distribution can be calculated and the Wurtz

distance computed. However, in a real prediction market it is usually not possible to know all the information possessed by the participants. Entropy is a parameter which can only give information on the precision on the market. However, it has the characteristic that it can be used at any time, even while the market is operating, by both the researcher and the practitioner.

Measure	Level	Measure of Precision or Accuracy?
Correct, Incorrect, Not-significant	Research: After stage Practice: Ex-post	Precision and Accuracy
Wurtz distance between result and AIA distribution	Research: After stage Practice: Generally not possible to calculate	Precision and Accuracy
Entropy	At any stage	Precision
Decisiveness	On a group of results	Precision
Correctness	On a group of results	Accuracy

Table 8. Summary of market measurement instruments.

Decisiveness and correctness will be used extensively in this work. Decisiveness gives a very valuable characterization of the precision of a market. A market with higher precision is able to make a prediction more often - a very valuable characteristic to the prediction market organizer. Correctness is the ability of the market to correctly predict once it has become decisive - or reached a statistically significant result. The prediction market organizer is keenly interested in both high decisiveness and correctness.

3.1.2 What are the Mechanisms Involved in the Information Assimilation Process?

As mentioned, the idea and understanding that markets can assimilate information date back to some of the earliest work on economics. Smith (1776) and Hayek (1945) wrote

of this and the concept was used to build the fundamental *Theory of Rational Expectations* and the *Efficient Market Hypothesis* - important pillars in modern economics. The ability of markets to assimilate information was empirically shown by Plott (1988). In the present study the process of assimilation will be described in terms of two components:

- The collection of diverse information - individuals using their own information make a judgment about the implications of that information and bring this to the market. This process is sometimes referred to as dissemination.
- The market's judgment on the implications of this gathered information - an inductive process which Plott (2003) and others call information aggregation.

A market without feedback can be studied to isolate the collection process. This process, if present, should enable a market to perform better than a simple collection of its components.

Hypothesis 1:

Markets without feedback can collect information and outperform the average of their parts.

Feedback markets can be compared to no-feedback markets to understand the effect of the inductive process. The level of agreement in the markets can be studied with the market entropy.

Hypothesis 2:

Feedback markets have an inductive process, above and beyond collection, which creates agreement which increases precision.

Feedback vs. no-feedback markets can be compared to estimate the amount of information which is being assimilated. The evaluation of the Wurtz distance between the cumulative allocation and the AIA distribution is a market measurement which characterizes the degree of information assimilation.

Hypothesis 3:

Feedback markets have an inductive process, above and beyond collection, which can assimilate more information than markets without feedback.

The information cascade effect which develops a mirage as described in Section 2.3.2.1 is expected to be seen on occasion in feedback markets. It is expected that this would be data and path dependent (e.g., a Polya process cf. Lloyd & Kotz, 1977). Various tests will be used to detect their presence in the data.

Hypothesis 4:

Information mirages are present in markets with feedback. They are an artifact of the inductive process. These mirages are an error mechanism which reduces the market's overall accuracy.

Table 9 presents a summary of research hypotheses for the research question "What are the mechanisms involved in the information assimilation process?"

Research Question	Hypothesis	Relevance
What are the mechanisms involved in the information assimilation process?	1. Markets without feedback can collect information and outperform the average of their parts.	How does a market achieve its results and how can it be managed?
	2. Feedback markets have an inductive process, above and beyond collection, which creates agreement which increases precision.	
	3. Feedback markets have an inductive process, above and beyond collection, which can assimilate more information than markets without feedback.	
	4. Information mirages are present in markets with feedback. They are an artifact of the inductive process. These mirages are an error mechanism which reduces the market's overall accuracy.	What are the error mechanisms in markets? And how can they be managed?

Table 9. Summary of research hypotheses for the first research question "What are the mechanisms involved in the information assimilation process?"

3.1.3 How does the Presence and Type of Feedback impact Market Performance?

The addition of feedback is expected to add the inductive process of information aggregation which should increase the precision of the market, but this process is also expected to create information mirages which will reduce accuracy.

Hypothesis 5:

Adding feedback increases the decisiveness of a market at the expense of correctness.

Three levels of feedback will be explored in the present study. First, no feedback on ticket allocations makes the market an incentivized poll. For feedback, the total accumulated tickets can be represented in two formats, as a frequency or as a percentage of the total. When total tickets numbers are shown, participants can derive the number of participants who have preceded them. However, if percentages of total tickets are shown then this will not be possible. In addition the percentage vs. frequency may impact the participant's estimation of Bayesian probabilities (Gigerenzer & Hoffrage, 1995).

Table 10 summarizes the feedback types of interest.

Factor	Level	Description
Feedback Type	No-feedback	No feedback is provided – incentivized poll
	Percentage allocation	Percentages of total accumulated allocated tickets shown
	Total Allocation (frequency)	Total accumulated allocated tickets shown

Table 10. Market Structure factors of Interest

Displaying frequency data on ticket allocations is expected to maximize the observational learning, which should result in increased decisiveness over percent feedback markets at the expense of some level of correctness.

Hypothesis 6:

The type of feedback can modulate the induction process - more informative feedback increases decisiveness at the cost of correctness.

Table 11 presents a summary of research hypotheses for the research question "How does the presence and type of feedback impact market performance?"

Research Question	Hypothesis	Relevance
How does the presence and type of feedback impact market performance?	5. Adding feedback increases the decisiveness of a market at the expense of correctness.	What is the relationship between prediction market designs and performance?
	6. The type of feedback can modulate the induction process - more informative feedback increases decisiveness at the cost of correctness.	Which structure should an organizational decision maker choose for their particular task?

Table 11. Summary of research hypotheses for the second research question "How does the presence and type of feedback impact market performance?"

3.1.4 How do Individuals use Information in the Allocation Process?

In the absence of feedback (that is, information on how the others participants are allocating their tickets) the research discussed in Section 2.2 suggests that many participants will act Bayesian-like in their allocations over a range of inputs. However, the deviation from Bayesian behavior is difficult to predict. For example, consider the case where a participant receives a ball draw of all the same type. Base rate neglect (section 2.2.2.2.1) and pseudodiagnosticity (section 2.2.2.2.3) would suggest the participant would overestimate the probability that the ball drawn is the correct state. However, conservatism (section 2.2.2.2.2) would suggest an underestimation of the probability.

Hypothesis 7:

In the absence of public data, participant's allocate tickets in a Bayesian-like method.

Information integration theory as described in section 2.2.2.2.1 would suggest that participants generally use algebraic rules to combine public and private information.

Hypothesis 8:

In the presence of public data, participants allocate tickets by integrating public and private data

Drawing from the discussion of preference reversals in section 2.2.3.2.3 it is expected that a large percentage of participants will make percentage bets (bets with a high probability of winning) , with a fewer number maximizing expected winnings. Or, in the terms used by Slovic and Lichtenstein (2006), they will favor P bets (probability) over \$ bets.

Hypothesis 9:

Participants will favor bets with a higher probability of winning over bets which maximize expected winnings

Finally, based on prior work, it is expected that there will be significant diversity in the overall individual strategies.

Research Question	Hypothesis	Relevance
How do individuals use information in the allocation process? What are the models for individual behavior?	7. In the absence of public data, participants allocate tickets allocate tickets in a Bayesian-like method	Can the performance of the market be improved by changing participants' behavior?
	8. In the presence of public data, participants allocate tickets by integrating public and private data	Who should be invited to participate?
	9. Participants will favor bets with a higher probability of winning over bets which maximize expected winnings	

Table 12. Summary of research hypotheses for the third research question "How do individuals use information in the allocation process?"

3.1.5 Summary of Research Hypotheses

Table 13 presents a summary of the research questions, associated hypotheses and relevance to the prediction market organizer.

Research Question	Hypothesis	Relevance
What are the mechanisms involved in the information assimilation process?	1. Markets without feedback can collect information and outperform the average of their parts.	How does a market achieve its results and how can it be managed?
	2. Feedback markets have an inductive process, above and beyond collection, which creates agreement (precision).	
	3. Feedback markets have an inductive process, above and beyond collection, which can assimilate more information than markets without feedback.	
	4. Information mirages are present in markets with feedback. They are an artifact of the inductive process. These mirages are an error mechanism which reduces the market's overall accuracy.	What are the error mechanisms in markets? And how can they be managed?
How does the presence and type of feedback impact market performance?	5. Adding feedback increases the decisiveness of a market at the expense of correctness.	What is the relationship between prediction market designs and performance? Which structure should an organizational decision maker choose for their particular task?
	6. The type of feedback can modulate the induction process - more informative feedback increases decisiveness at the cost of correctness.	
How do individuals use information in the allocation process? What are the models for individual behavior?	7. In the absence of public data, participants allocate tickets allocate tickets in a Bayesian-like method	Can the performance of the market be improved by changing participant's behavior? Who should be invited to participate?
	8. In the presence of public data, participants allocate tickets by integrating public and private data	
	9. Participants will favor bets with a higher probability of winning over bets which maximize expected winnings	

Table 13. Summary of research questions, associated hypotheses and relevance to the prediction market organizer.

3.2 Experiment Design

A simplified version of the experimental prediction markets used by Plott et al. (2003) was used. With a few modifications, this market can be conducted asynchronously via an online survey.

The overall design is shown in Table 14. Feedback type has three levels: no-feedback (called here an incentivized poll); total accumulated tickets shown as frequency; or, total accumulated tickets shown as percentages of the total.

		Market Symbol
Feedback Type	No-feedback	IP (Incentivized Poll)
	Percentage of ticket allocations show (%)	Pct (or %)
	Total ticket Allocation Shown (Frequency)	Freq

Table 14. Overview of experimental design

A flow chart of the survey is shown in Figure 3. Participants were recruited from Amazon Mechanical Turk. After consent and a short set of demographic questions they participated in two incentivized poll markets. Next, they were randomly assigned to one of the feedback groups – either being shown total accumulated tickets as a frequency or as a percentage of the total. After completing these four markets they completed a short set of survey questions.

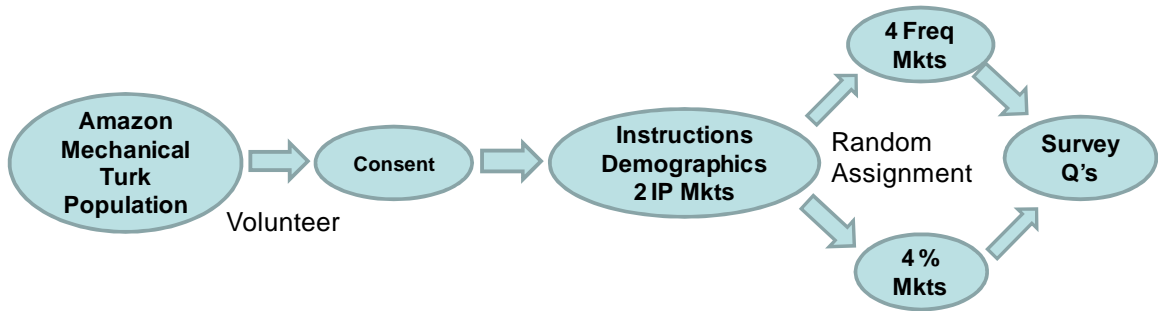


Figure 3. Experiment flow chart

Two experiments were conducted. The primary difference was the draw tables used.

Draw tables can be produced prior to the market implementation and individual draw sets taken sequentially from them.

3.2.1 Simplified Experimental Prediction Market

The market used in the present research is a simplified version of the prediction market studied by Plott et al. (2003). Participants will be given private data and, depending on the market structure, may be able to observe how other participants have acted (public data). The private data consists of a set of draws from an urn. As used by Plott et al. (2003), the urn contains 15 balls with a distribution biased by a prior selection of the correct state.

Figure 4 illustrates the operation of a frequency market. To begin, the correct state of nature, that which participants are attempting to ascertain, is chosen at random from an urn (or bingo cage) consisting of six balls – here considered simply to be letters A through F. Suppose this random selection chooses B to be the correct state. The second urn (or bingo cage) would then be populated with 5 B balls, and 2 balls of each of

the other types. Each market participant is then given the results of three draws, with replacement, from the bingo cage. For example, from the figure the first set of three draws was E, B and F. This set of three draws will be called here a draw set. These draw sets are the private information of the participants. Each participant was given an allocation of 100 tickets to distribute across the six states. The participant's task is to allocate these tickets in the best way given their particular set of information. As the market begins, the first participant is given his private information (draw set) of E,B,F. He sees the prior allocation is all 0 and can understand that he is the first market participant. For this example, imagine he allocates his 100 tickets as 12 to state A, 8 to state B and C, 16 to state D and E and 40 to state F. The next participant is shown a draw set of C,D,A and the public information, the total accumulated ticket allocation by prior respondents, of 12,8,8,16,16,40 (the values allocated to states A through F respectively). They would then allocate their 100 tickets based on this information. In this illustration they allocated tickets as 24,10,23,23,10,10 to states A through F. The total accumulated number of tickets becomes 36,18,31,39,26,50. These are the totals which would be shown to participant 3 along with their draw set of F,D,E.

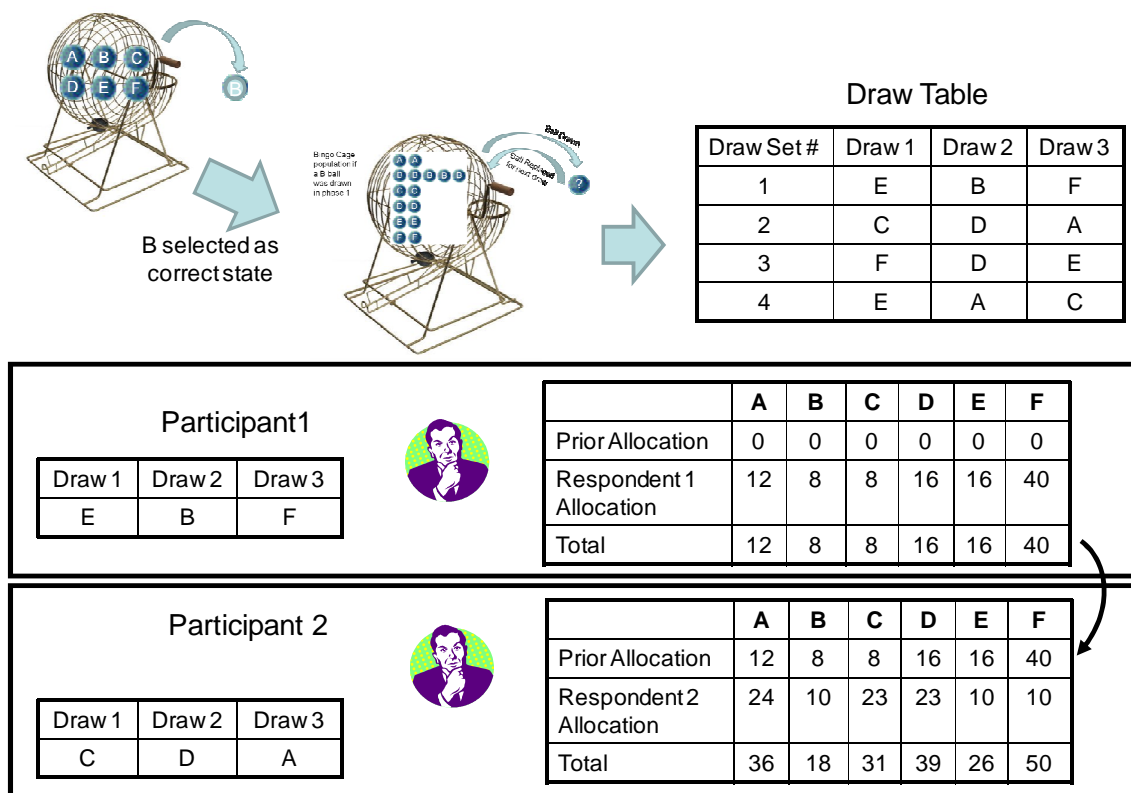


Figure 4. Illustration of market operation

This market process places two restrictions not seen in the markets used by Plott et al. (2003) in order to operate asynchronously:

- Each participant is given a set order and must allocate all their tickets at that time. Generally, participants would prefer to place their tickets later in the process. However, if the scope is a series of markets, and the participant's position in the allocation queue is randomly chosen, then, on average, no participant would be disadvantaged over the series of markets.
- Each participant must allocate all of their tickets. While not a requirement of the markets used by Plott et al. (2003), it is conceptually consistent with these markets since there was no benefit to holding unallocated tickets at the end of the market since they did not carry over to the next market.

Rather than assembling all participants together, the modified structure allows participants to respond asynchronously – thereby greatly boosting the ability to gather large numbers of responses. The asynchronous nature of the proposed structure may

also have value as an organizational market – allowing organizations to collect data from busy participants who may be geographically dispersed across time zones.

3.2.2 Random Draw Experiment

For the random draws experiment, hundreds of draw tables were produced with Matlab's rand command which uses the Ziggurat algorithm (Marsaglia & Tsang, 2000). These draw tables were then used sequentially as the experiment progressed. The same draw tables were applied to each market type (no-feedback, frequency feedback and percent feedback).

3.2.3 Mirage Prone Draw Experiment

In the mirage prone draw set experiment, one draw table was reused for all markets.

To select the draw table a large number of possible draw tables were generated. These draw tables were processed through a simulator (see section 5.2) to analyze the potential performance of each draw table in different prediction markets. Table 15 shows the draw table to be used in the experiment. Here, A is the correct state. For example, the first participant would receive a draw set of D, F, D. The second participant would receive a draw set of E,F,D.

Correct State = A			
Draw Set	1st draw	2nd draw	3rd draw
1	D	F	D
2	E	F	D
3	F	C	A
4	D	D	E
5	A	D	E
6	A	E	B
7	A	A	B
8	E	C	C
9	A	A	B
10	C	F	B
11	A	B	A
12	C	B	D
13	A	A	A

Table 15. Experimental Draw Table for Mirage Prone Draw Experiment.

This particular draw table was chosen due to its ability to generate mirages in the simulator environment. Table 16 shows the statistics of the draw table. Compared to the urn, the signal for the correct state is weaker (25% vs. 33.3% for the correct state in the case of 12 draw sets). Furthermore, the draws for the correct state come later in the draw table. While the simulator showed this has no effect in a market without feedback, the simulator demonstrated that this arrangement is capable of producing information mirages in feedback markets.

State	Draw Table			Urn	
	Count	Percent		Count	Percent
A	9	25.0%		5	33.3%
B	6	16.7%		2	13.3%
C	5	13.9%		2	13.3%
D	7	19.4%		2	13.3%
E	5	13.9%		2	13.3%
F	4	11.1%		2	13.3%

Table 16. Statistics of the draw table compared to the Urn.

The mirage tendency of the draw set can be illustrated in Table 17 and Table 18. Table 17 shows the draw sets at each stage (participant) as well as the AIA posterior probability given the cumulative draws. Note that for stages one through eight, D is the most likely state (largest AIA probability). From stages 9-13, state A, which is the urn correct state, becomes the most likely. Table 18 illustrates the cumulative allocation if participants had access to all accumulated draws and they allocated their 100 tickets in direct proportion to the AIA posterior probability. Even with these rational responses, an information mirage would develop on state D.

				AIA Posterior Probability					
Random Draw									
1st draw	2nd draw	3rd draw	Stage	A	B	C	D	E	F
D	F	D	1	7.8%	7.8%	7.8%	49.0%	7.8%	19.6%
E	F	D	2	3.7%	3.7%	3.7%	57.1%	9.1%	22.8%
F	C	A	3	6.3%	2.5%	6.3%	39.3%	6.3%	39.3%
D	D	E	4	2.0%	0.8%	2.0%	77.8%	5.0%	12.4%
A	D	E	5	2.2%	0.4%	0.9%	85.6%	5.5%	5.5%
A	E	B	6	4.9%	0.8%	0.8%	76.4%	12.2%	4.9%
A	A	B	7	24.1%	1.5%	0.6%	60.2%	9.6%	3.9%
E	C	C	8	20.5%	1.3%	3.3%	51.2%	20.5%	3.3%
A	A	B	9	61.1%	1.6%	1.6%	24.4%	9.8%	1.6%
C	F	B	10	57.1%	3.7%	3.7%	22.8%	9.1%	3.7%
A	B	A	11	88.1%	2.3%	0.9%	5.6%	2.3%	0.9%
C	B	D	12	77.8%	5.0%	2.0%	12.4%	2.0%	0.8%
A	A	A	13	98.2%	0.4%	0.2%	1.0%	0.2%	0.1%

Table 17. Mirage prone draw set and AIA posterior probability.

			Accumulated Allocation			
Stage	A	B	C	D	E	F
1	8	8	8	49	8	20
2	11	11	11	106	17	42
3	18	14	18	145	23	82
4	20	15	20	223	28	94
5	22	15	21	309	34	100
6	27	16	21	385	46	105
7	51	17	22	445	56	108
8	71	19	25	497	76	112
9	133	20	27	521	86	113
10	190	24	31	544	95	117
11	278	26	31	550	97	118
12	355	31	33	562	99	119
13	454	32	34	563	99	119

Table 18. Accumulated allocation if each participant allocates in proportion to the current AIA probability.

3.3 Survey Implementation

In order to generate a large sample of participants the Amazon Mechanical Turk (<https://www.mturk.com/mturk/welcome>) on-line recruitment mechanism was used.

Amazon Mechanical Turk (referred to as AMT) was developed as a way to recruit workers for small tasks (called by AMT as HITs for Human Interface Tasks). For example, businesses often use AMT workers (referred to as Turkers) to screen customer blog posts for inappropriate comments. Social scientists are finding the internet (Birnbaum, 2004) and AMT (Mason & Suri, 2010) to be an excellent source of high quality participants.

The interaction between participants necessitated a more complex survey implementation than normally seen in social science experiments. Figure 5 illustrates the survey architecture. AMT workers log into their accounts at Amazon.com and access AMT

tasks (HITs) with their web browser. The entire survey ran within the AMT frame and had to interface into the AMT account information. The survey itself was hosted on a separate site (lunarpages.com). The php scripts ran on these hosted servers and accessed a MySQL database to store information and retrieve summaries of responses from prior participants. All interfaces to AMT account settings, survey scripts and the MySQL database were available via the internet. The final set of scripts contained over 400 lines of code.

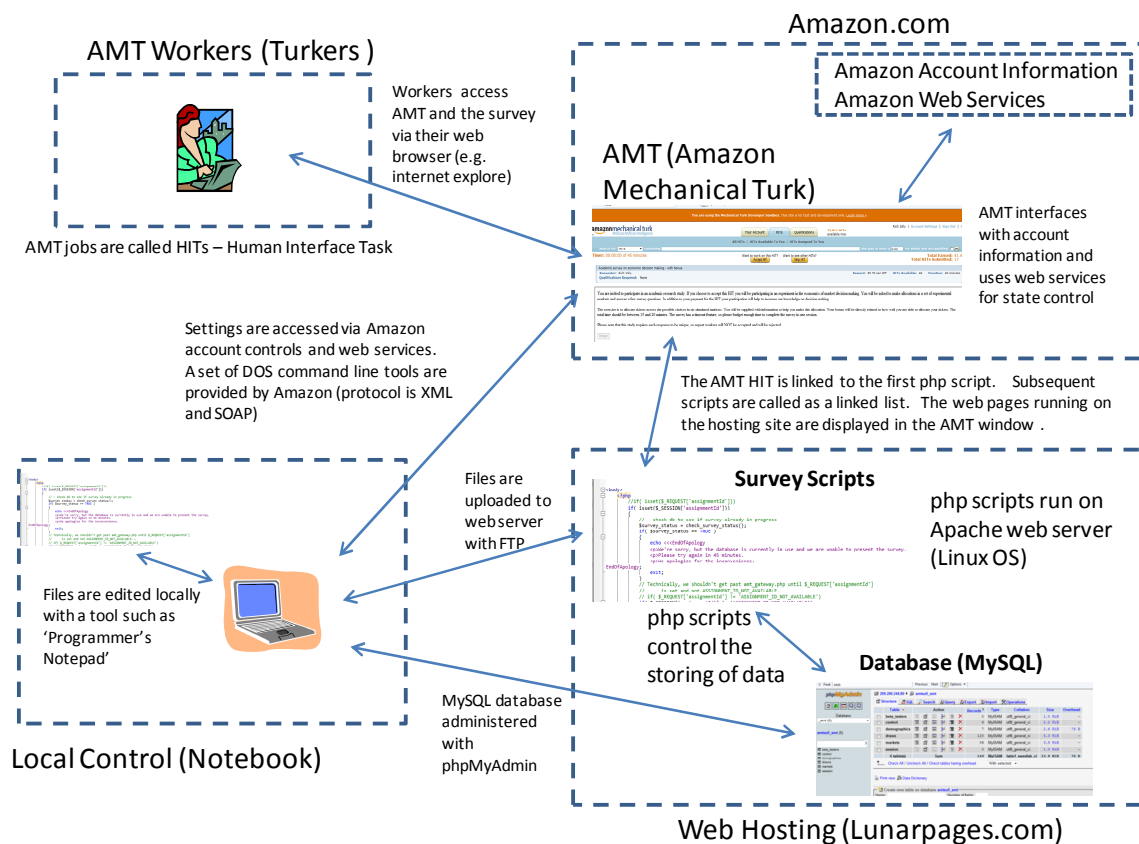


Figure 5. Survey Architecture

The survey includes (see Appendix for details):

- An IRB approved consent form.
- Collection of basic demographics: gender, age, education level and major

- An attention test - a Likert scale question which simply asked 'select strongly agree'
- A set of instructions on the incentivized poll market
- Two no feedback (IP) markets
- For the random draws experiment a third IP market was given which had a fixed draw and was used for screening purposes (see section 4.1.4)
- A set of instructions for the feedback market (either frequency or percent depending on the random assignment).
- Four feedback markets (Figure 6 gives a screen shot of the actual implementation)
- A post test survey question "Did you encounter a situation where there was a contradiction between your private (draw) information and the market information?"
- If the participant responded yes, they were presented with the question - ' How did you resolve the contradiction?' with choices of:
 - I relied on private information
 - I relied on public information
 - I attempted to integrate the two information sources
 - I just spread out my tickets across all options

First Draw	Second Draw	Third Draw
C	E	C

State	A	B	C	D	E	F
Tickets Allocated	0	0	0	0	0	0

Tickets Remaining to Allocate	100
-------------------------------	-----

State	A	B	C	D	E	F	Type your allocation into the shaded boxes
Tickets Allocated	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	<input type="text" value="0"/>	

Submit Reset

Only 3 market(s) left. Remember you must complete the *entire* survey to be eligible for payment.

Figure 6. Actual implementation of the frequency feedback market.

The markets were designed to stagger the start. This was done so that each participant would see a variety of total accumulated public data in the feedback markets and a range of private data in the mirage prone draw experiment. To illustrate, if there were no staggering, the first participant would see a total accumulated allocation of tickets as 0 for each state in each of the four feedback markets. By contrast, the last participant would

have the advantage of seeing the allocations of all prior participants for each market. By staging the start, it allowed each participant to see a range of total accumulated responses. This not only makes the potential rewards fairer, but also improves the diversity of responses obtained. This staging will be explained in the following sections

3.3.1 Staging in Random Draw Experiment

To stage, or stagger, the start, a number of seed responses were generated and entered into the MySQL database. Table 19 illustrates the staging method used. To explain, consider the case of participant 1. After completing the no feedback markets they would receive draw set 11 in their first feedback market. As public data, they would see the accumulated allocation of seed responses 1 through 10. For their second market, they would be the first participant, receiving draw set 1 and would be shown a total accumulated allocation of 0 for each state as public information. In their third feedback market they would receive draw set 16 and see the accumulated allocation of seeds 1 through 15 as the public data. Finally, in market 4 they would receive draw set 6 and see the total accumulated allocation of seeds 1 through 5. In this example, the participant's total accumulated allocation ranged from 0 in market 2 to 1,500 in market 3. All markets which included seeded values were excluded from the analysis. So, there is a research cost for this staging.

Draw Set #	Feedback Market #1	Feedback Market #2	Feedback Market #3	Feedback Market #4
1	Seed 1	Participant 1	Seed 1	Seed 1
2	Seed 2	Participant 2	Seed 2	Seed 2
3	Seed 3	Participant 3	Seed 3	Seed 3
4	Seed 4	Participant 4	Seed 4	Seed 4
5	Seed 5	Participant 5	Seed 5	Seed 5
6	Seed 6	Participant 6	Seed 6	Participant 1
7	Seed 7	Participant 7	Seed 7	Participant 2
8	Seed 8	Participant 8	Seed 8	Participant 3
9	Seed 9	Participant 9	Seed 9	Participant 4
10	Seed 10	Participant 10	Seed 10	Participant 5
11	Participant 1	Participant 11	Seed 11	Participant 6
12	Participant 2	Participant 12	Seed 12	Participant 7
13	Participant 3	Participant 13	Seed 13	Participant 8
14	Participant 4	Participant 14	Seed 14	Participant 9
15	Participant 5	Participant 15	Seed 15	Participant 10
16	Participant 6	Participant 16	Participant 1	Participant 11
17	Participant 7	Participant 17	Participant 2	Participant 12
18	Participant 8	Participant 18	Participant 3	Participant 13
19	Participant 9	Participant 19	Participant 4	Participant 14
20	Participant 10	Participant 20	Participant 5	Participant 15

Table 19. Staging for random draws experiment.

3.3.2 Staging in Mirage Prone Draw Experiment

Six versions of the mirage prone draw table (Table 15) were used. The 5 additional versions are simple 1:1 mappings from the original as shown in Figure 7. Each original state is mapped to a new state. This technique allows the correct state to vary and provides more variety of draw sets while still allowing comparison of results.

Original Draw State	Mapped State				
	Mapping 1	Mapping 2	Mapping 3	Mapping 4	Mapping 5
A	C	D	B	D	E
B	F	B	F	E	A
C	A	A	E	F	F
D	B	C	C	A	B
E	D	F	A	C	D
F	E	E	D	B	C

Figure 7. Mappings used to create alternate version of draw table

Table 20 and Table 21 illustrate the staging method for the mirage prone experiment. IPA-1 and IPB-1 are the first two IP markets. FA-1 through FD-1 are the first four frequency feedback markets. PA-1 through PD-1 are the first four percentage feedback markets. IPB-1, FB-1, FC-1, FD-1, PB-1, PC-1 and PD-1 were all seeded with computer generated responses and are indicated with light blue shading in Table 20. To understand this consider a series of participants - labeled in the table as Resp1 for the first participant, Resp2 for the second responded, and so forth. A random assignment into either the frequency or percentage markets is simulated. For example, participant 1 (resp1) would first receive draw set #1 (that is be the first participant in market IPA-1) from transform 1 (see Figure 7 - draw set #1 from transform 1 is 2,5,2 or B,E,B) for his first IP market. In his second IP market he would receive the thirteenth draw set from the untransformed draw table (1,1,1 or A,A,A). I illustrate that the random assignment would place him into the frequency markets. In his first frequency market he would receive draw set number 1 from transform 2 of the draw table (2,5,2 or B,E,B). Since he is the first participant in this market his public data will be all zeros (0,0,0,0,0,0). In his next frequency market, FB-1, he will receive the fourth draw set from transformation

draw table 3 (2,2,1 or B,B,A). The public data he will see is the sum of the seed allocations SB1-1 + SB1-2 + SB1-3 (all computer generated). In this way, each participant in the feedback markets will have a variety in the number of prior participants. Note that if the seed allocations were not used, participant 1 would be the first participant in every market. Every draw he saw would be simple transformations of the same draw (two balls the same). In every feedback market he would see no public information. Also, without the seeding and staging, participant 13 would see all ball draw sets as three balls the same (e.g., A,A,A) and he would have access to the allocations of all twelve prior participants as his public information. Finally, any market with seed values are not used in the calculation of statistics for market performance (that is, they are removed from the sample).

		IPA-1	IPB-1		FA-1	FB-1	FC-1	FD-1
Transform		1	none		2	3	4	5
Correct State		3	1		4	2	4	5
	1	Resp 1	NA		Resp 1	SB1-1	SC1-1	SD1-1
	2	Resp 2	NA		Resp 3	SB1-2	SC1-2	SD1-2
	3	Resp 3	NA		Resp 5	SB1-3	SC1-3	SD1-3
	4	Resp 4	NA		Resp 7	Resp 1	SC1-4	SD1-4
	5	Resp 5	NA		Resp 10	Resp 3	SC1-5	SD1-5
	6	Resp 6	NA		Resp 15	Resp 5	SC1-6	SD1-6
	7	Resp 7	NA		Resp 17	Resp 7	Resp 1	SD1-7
	8	Resp 8	NA		Resp 18	Resp 10	Resp 3	SD1-8
	9	Resp 9	NA		Resp 19	Resp 15	Resp 5	SD1-9
	10	Resp 10	NA		Resp 21	Resp 17	Resp 7	Resp 1
	11	Resp 11	NA		Resp 22	Resp 18	Resp 10	Resp 3
	12	Resp 12	NA		Resp 26	Resp 19	Resp 15	Resp 5
	13	Resp 13	Resp 1		Resp 27	Resp 21	Resp 17	Resp 7

Table 20. Illustration of the staging method for the mirage prone experiment- IP and Frequency Markets.

		PA-1	PB-1	PC-1	PD-1
Transform		2	3	4	5
Correct State		4	2	4	5
	1	Resp 2	SB1-1	SC1-1	SD1-1
	2	Resp 4	SB1-2	SC1-2	SD1-2
	3	Resp 6	SB1-3	SC1-3	SD1-3
	4	Resp 8	Resp 2	SC1-4	SD1-4
	5	Resp 9	Resp 4	SC1-5	SD1-5
	6	Resp 11	Resp 6	SC1-6	SD1-6
	7	Resp 12	Resp 8	Resp 2	SD1-7
	8	Resp 13	Resp 9	Resp 4	SD1-8
	9	Resp 14	Resp 11	Resp 6	SD1-9
	10	Resp 16	Resp 12	Resp 8	Resp 2
	11	Resp 20	Resp 13	Resp 9	Resp 4
	12	Resp 23	Resp 14	Resp 11	Resp 6
	13	Resp 24	Resp 16	Resp 12	Resp 8

Table 21. Illustration of the staging method for the mirage prone experiment - Percent Markets.

3.4 Research Analysis Overview and Tools

In addition to those measures discussed in section 3.1.1 several other tools and techniques are used in the analysis.

3.4.1 Evaluating Draw Sets

To evaluate the effect of different draw sets a quantitative method of evaluating the draw set is needed. As a draw set progresses through the draws the AIA value will eventually signal the correct Urn state with a high probability. A Wurtz distance can be calculated between the current AIA and the final AIA. Evaluating this as a time series can be a valuable way to study and characterize draw sets. Consider the draw set illustrated in Figure 8. This plot shows the Wurtz distance between the current AIA and the final AIA (after 20 draws) at each stage (participant) in the draw set. In this example the correct

state is F. Many early draws of F, quickly drive the AIA to a high probability for state F. By draw number 7, the probability that the correct state is F has reached nearly 100%.

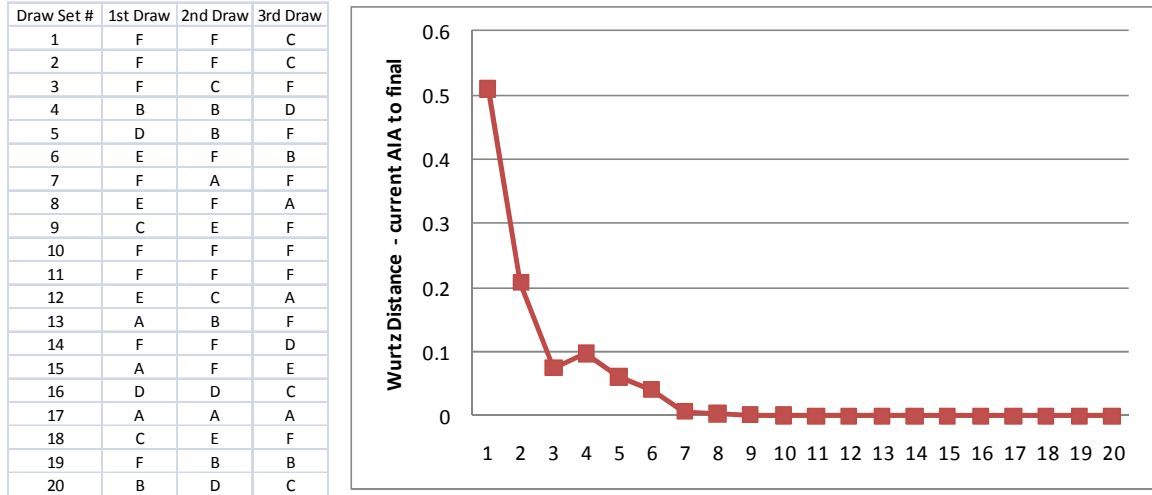


Figure 8. A well behaved draw table. The x axis of the graph is the market stage (or number of participants).

Contrast this to the draw set illustrated in Figure 9. The correct state is E. It takes nearly 13 draws before the correct state is predicted with greater than 80% probability. It is anticipated that the shape of the path towards the asymptotic value of AIA will influence feedback prediction markets. A simple quantitative metric which could be used to characterize this aspect of the draw set is simply the sum of these AIA Wurtz distances as shown in the equation below. This can be thought of as a measure of the area under the curves. The sum of the Wurtz distances for the draw set in Figure 8 is 1.00 while the sum for the draw set in Figure 9 is 7.29.

$$\text{Sum of AIA distances} = \sum_{i=1}^N \text{Wurtz Distance}(AIA_i - AIA_N)$$

Draw Set #	1st Draw	2nd Draw	3rd Draw
1	E	E	F
2	F	A	F
3	B	E	D
4	E	F	F
5	E	D	D
6	D	F	D
7	D	A	B
8	D	F	B
9	E	F	E
10	E	E	E
11	C	D	D
12	A	C	E
13	E	E	D
14	C	A	E
15	E	C	B
16	B	F	E
17	A	E	B
18	F	D	A
19	F	D	E
20	F	C	E

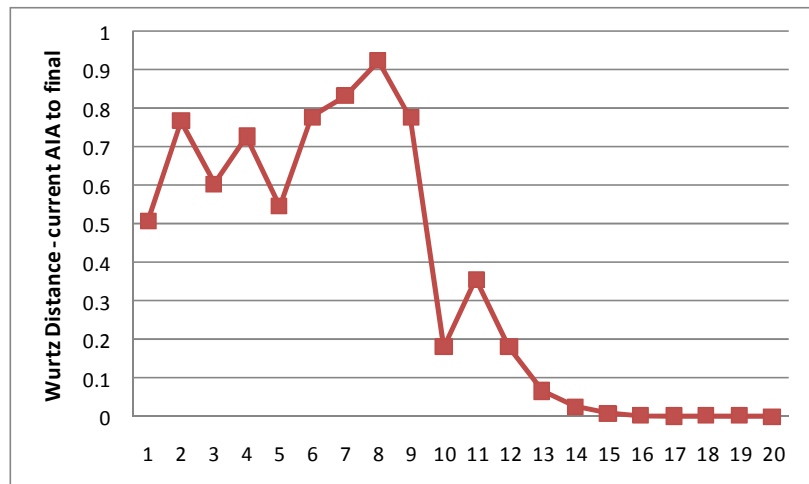


Figure 9. A non-well behaved draw table. Many draws are required to reach the AIA asymptote
The x axis of the graph is the market stage (or number of participants).

3.4.2 Understanding Individual Allocations

Individual responses can be compared to hypothetical responses to be categorized. This can lead to insights into the market processes, be useful for screening purposes and be useful for building models of markets as explored in the discussion section.

There are various theoretical frameworks available when considering different methodologies for allocating tickets. The framework shown in Figure 10 compares methods based on their use of information and computation required. MaxiMin (maximize the minimum) in this market uses no data and requires the minimum level of computation - simply allocating the same number of tickets across each state. A percent allocation method would allocate tickets in direct proportion to their probabilities. This would require more computation to determine these percentages. This task would get more burdensome if public data is available and were used in the allocation determination. Finally, MaxEV (maximize the expected value) requires one calculation step beyond Percent. That is, it must allocate to maximize winnings given the estimated probabilities.

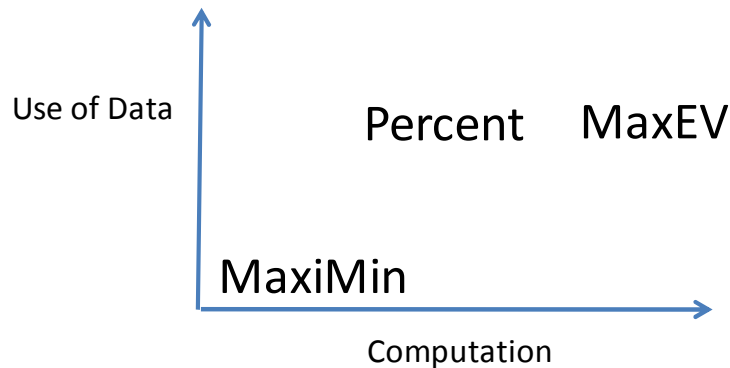


Figure 10. Qualitative comparison of basic allocation methods

First consider the IP markets where the only information available is the draw set. Two methods of computing probabilities were considered. First, probabilities could be estimated by the simple percentages from the ball draw. For example, if the draw were A,A,C a participant might estimate the probability to be 66.6% for A and 33.3% for C and then allocate tickets in that ratio. An alternative would be to calculate the Bayesian probability. Given the probabilities were equal for all states when the correct state was initially selected, one can determine the posterior probabilities of any given ball being correct given a type of draw set. Table 22 shows the calculated Bayesian posterior probabilities given any set of ball draws. A participant could then allocate tickets in these proportions. The expected value of an allocation can be calculated by multiplying the number of tickets allocated on any state by the probability of that state and summing all the products. It is straightforward to show that to maximize the expected value one should allocate all tickets on the option with the highest probability. In the case where there are options with equal probabilities, any allocation where all tickets are distributed across these equal probability options will be a maximum.

	Most Frequent (e.g. A)	2 nd Most Frequent (e.g. B)	2 nd Most Frequent (e.g. C)	Each Not Drawn (e.g. DEF)
Three matching draws (e.g. AAA)	75.8%	4.8%	4.8%	4.8%
Two matching draws (e.g. AAB)	49.0%	19.6%	7.8%	7.8%
Three different draws (e.g. ABC)	23.8%	23.8%	23.8%	9.5%

Table 22. Bayesian probabilities given the draw sets shown.

Table 23 shows the result of categorizing the IP market responses of those who passed the attention test from the mirage prone draw set experiment. MaxiMin was the most prevalent strategy followed closely by allocation to Bayesian probabilities. Allocating to simple percentages was quite rare. As discussed in the literature review, this is consistent with prior studies. To simplify further work, the style of allocating to simple percentages was dropped.

	Count	Percent
Match allocation to simple percentage from ball draw	14	3.7%
Match allocation to Bayesian probabilities	158	41.9%
Maximize expected winnings based on probabilities	27	7.2%
Equal allocation on each state (MaxiMin)	178	47.2%

Table 23. Categorization of responses in IP market to theoretical models from the random draws experiment. In this analysis, only participants who passed the attention test are considered

Extending the calculation of probabilities to the use of both public and private data is complicated. Two methods were used in the present study as illustrated in Figure 11. In this example the participant received a draw set of A,A,C and was shown a public distribution of accumulated tickets of 77,51,171,51,51,299. Two interpretations of the

public data were considered here. First, the simple percentages could be calculated as shown (e.g., 77 tickets allocated to state A is $77/700$ or 11%). This is all the information available to participants in the percentage markets. In a second interpretation a participant could attempt to infer the draw sets of prior market participants based on how they allocated their tickets. Some assumption would need to be made on how participants had allocated tickets in order to use this method. The method used here assumes each participant allocated tickets in direct proportion to their draw. In this example, 700 total tickets have been allocated indicating there were 7 prior participants (and thus 7 draw sets or 21 total balls drawn from the urn). The 51 tickets allocated to state B could imply that there have been 1.53 ($51/700 * 21$) balls drawn for that state. These assumptions could then be used in a calculation of Bayesian probabilities. In the case where public information is simply converted into percentages, this percentage can be used as the prior probability and the ball draw used to calculate the posterior probability. In the case where public ticket allocations are converted to some equivalent number of ball draws, then the method of calculating Bayesian probabilities discussed above for IP markets can simply be extended to include the private draw set combined with the ball draws as implied by the public data. In the example of Figure 11 these two methods arrive at different predictions for the most likely correct state. When using the public data as a simple percent, the private draw set of 2 balls on state A weighs the final probability more than the 42.7% prior probability of state F and predicts state A as the most likely with a 35.4% probability. However, if ticket counts are converted to ball draws based on simple allocation percents, the combined private draw set and inferred public draws imply state F is much more likely at 91.7% probability. The

method of converting the public allocation into ball draws can be considered to use more information available since it considers the total number of participants.

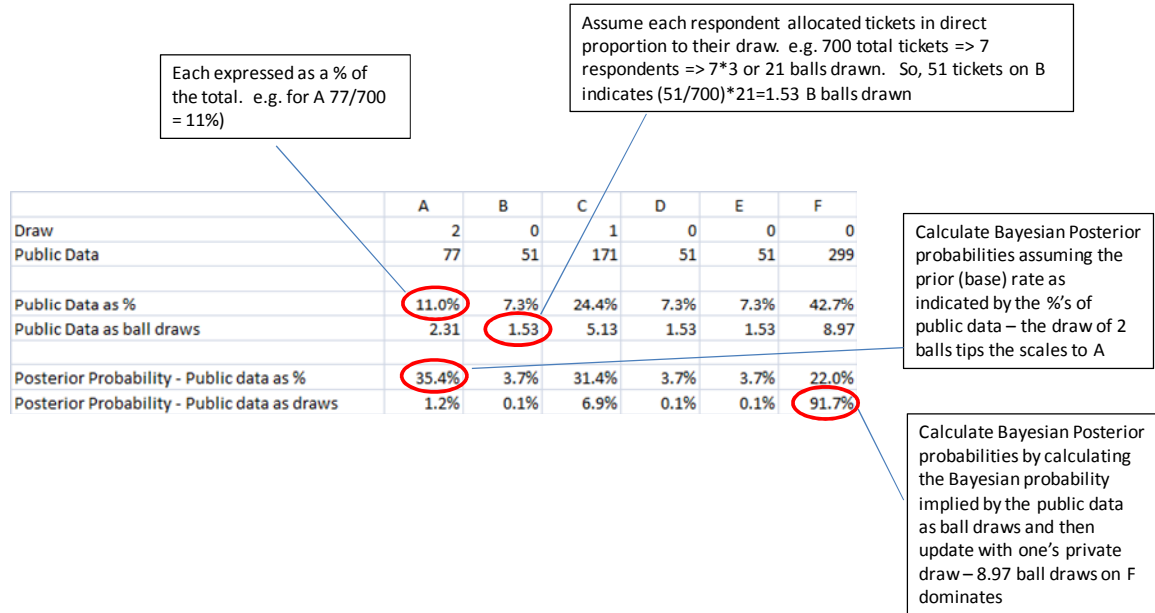


Figure 11. Possible methods of using Bayesian probability to combine public and private data.

Figure 12 shows a summary of the allocation styles used to characterize the responses in this experiment. This figure shows the predicted allocations given the private and public information shown at the top of the figure (that is a draw set of A,D,A and public information of 23,43,19,26,53,36). In addition, it compares the algorithm's predicted allocation to a hypothetical participant's allocation of 10,20,10,15,25,20 using the Wurtz distance. A modified MaxiMin implementation was used because it is not possible to evenly divide 100 tickets into 6 buckets. The modified MaxiMin algorithm used here allocates the extra 4 tickets across states with a ball in the draw set (even in situations with Public information). The next two algorithms use only private data (the draw set). First DrawPct calculates the Bayesian posterior probability based on the draw set and

allocates tickets accordingly. The DrawEV algorithm also calculates the Bayesian probability but then will maximize the expected value of the allocation by placing all tickets on the most likely correct state. Next, an algorithm that uses only public information is considered. The Mimic algorithm allocates tickets in direct proportion to the public allocation. The PublicEV algorithm will maximize the expected value assuming the probabilities implied by the public distribution of tickets. The final four algorithms integrate public and private data as described in the last paragraph and illustrated in Figure 12. AllPct uses the public allocation percentages as the prior in a Bayesian posterior probability calculation. AllPctEV then maximizes the expected value of the ticket allocations based on those calculated probabilities. AllDraws decomposes the public data and calculates an implied ball draw assuming prior participants had allocated tickets in direct proportion to their draw sets. AllDraws then calculates a Bayesian posterior probability by combining the private draw set with this implied public set of draws and then allocates tickets in proportion to the implied probabilities. The AllDrawsEV algorithm will allocate tickets to maximize expected value given the probabilities calculated in AllDraws.

Figure 12 shows that for this illustration the best fit to the participant's allocation is the Mimic algorithm as determined by having the smallest Wurtz distance. In general, the algorithms lower on this figure use more available information. Only the top three algorithms are used to categorize participant's allocations in IP markets. Only the top 7 are used to categorize participant's allocations in Percent markets. And, all are used to categorize participant's allocations in Frequency markets.

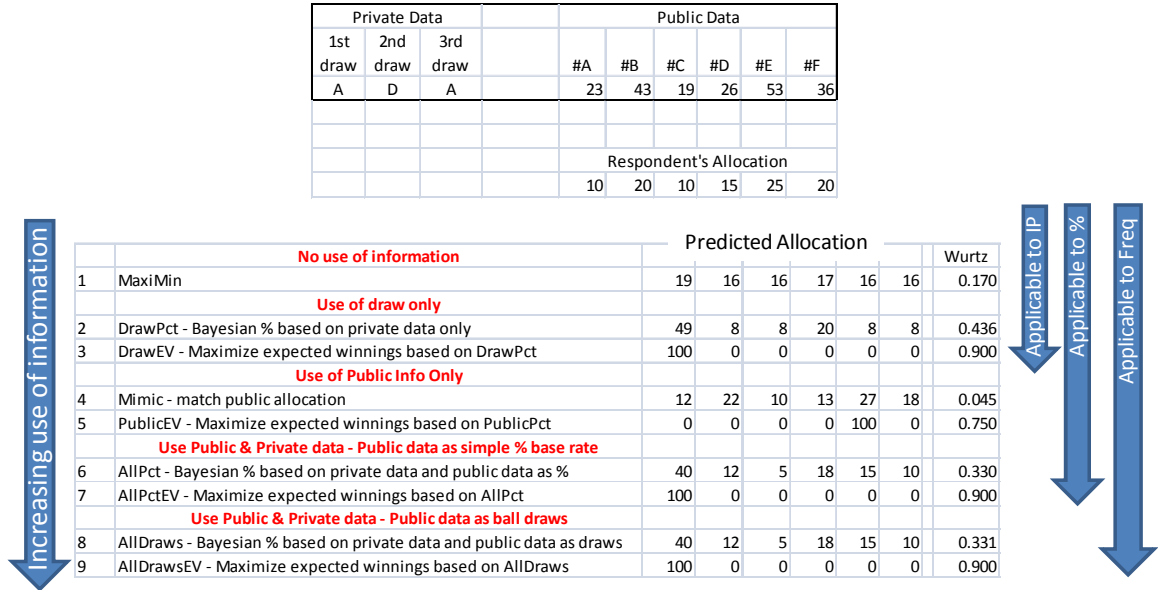


Figure 12. Summary of the allocation styles studied.

3.4.3 Inferential Tools and Considerations

Non-parametric inferential tools were generally used. Two by two contingency tables were evaluated with Fisher's Exact test (Fisher, 1954) due to its superior performance when entries are small. Differences in means were generally tested using Mann–Whitney–Wilcoxon U test (Mann & Whitney, 1947) to avoid any issues with the assumptions of underlying distribution characteristics. Generally, inferential tests assume independence between observations. Due to the nature of the experimental procedure, each participant participated in multiple markets. Due to the staggered start, no two markets had the exact same set of participants. However, there was some amount of overlap. While this is not a major challenge to the significance test results, this overlap produces some weakening of the independence assumption of most inferential

tests. The researcher is unaware of any correction for this slight bias, so readers should keep in mind that results may not quite as strong as the p-value suggests.

3.5 Potential Limitations of Experimental Methods

The key limitations are discussed.

3.5.1 Sequencing Participants

In the experiments of Plott et al. (2003) participants were free to allocate any number of tickets at any time. If a participant believes they can make better predictions using the public data, then there is incentive to wait. If a participant feels they have very good private data (e.g., a private draw of three balls all of the same type) they are also incented to wait since their allocation may influence others to allocate their tickets to the same bucket and thereby reducing their winnings (in parimutuel markets used by Plott but not in the fixed winning market).

The fixed winning market has slightly different dynamics. As in the Plott experiments there is incentive to see other's allocations in the sense that it reveals more private information. However, since each ticket has a fixed winning amount, independent of what other market participants have done, there's no incentive to wait to reveal your own private data to the group.

However, the proposed asynchronous market may have benefits to organizations and be a preferred method of conducting prediction markets. This asynchronous nature could be

valuable in a busy organization where it may be difficult to schedule a common time block across all market participants. Secondly, the waiting behavior described above, while benefitting the individual, may not benefit the organization. If this type of mechanism were used in an organization, the participant's turn could be determined with a random draw. As markets may be an ongoing event (e.g., a recurring forecast) then participants would be expected to get turns at all stages of the market (early to late).

3.5.2 Learning

Learning in games is an important consideration (Erev & Roth, 1998). This includes both learning in terms of the game itself, but also in how the participants act. Camerer (1992) has a good discussion on the topic relevant to markets:

Learning effects are mixed too. There is rapid learning in some simple market environments (e.g., in double auctions for commodities, Smith, 1982) but not in more complicated environments (Smith, Suchanek, & Williams, 1988; Camerer & Weigelt, 1990; Ball, Bazerman, & Carroll, in press). Furthermore, many studies suggest that experience or expertise are no guarantee of superior performance in naturally occurring tasks (e.g., Northcraft & Neale, 1987; Camerer & Johnson, 1991).

4 Experimental Results

The experimental results are reviewed. Some limited discussion is included in this section as well. The complexity of analysis requires that some of the discussion be provided with the results rather than later in the general discussion chapter.

4.1 General Considerations

This section will address some of the general considerations around the survey results.

4.1.1 Data Collection

Table 24 shows the data collection statistics.

The Random Draws Experiment was conducted from July 11, 2011 to July 27, 2011. 503 participants were used in the analysis. Participants earned a base of \$0.10 and a bonus based on their performance. The bonus was calculated as their total number of correct tickets divided by their total tickets times \$1.50 (the mean bonus amount was \$0.108). The Attention test (section 3.3) and no feedback screener (section 4.1.4) were used in this experiment.

The Mirage Prone Draw Experiment was conducted over a 5 day period from November 9, 2010 to November 13, 2010. During that time 316 participants entered the survey and 228 completed the survey. Each participant that finished was paid \$0.15 plus a bonus based on performance. The bonus was calculated as their total number of correct

tickets divided by their total tickets times \$1. The survey generated over 51k SQL records.

	Random draws experiment	Mirage Prone Data
Data Collection Period	7/11/11 to 7/27/11	11/9/10 to 11/13/10
Incentives	Base of \$0.10 plus bonus with a maximum of \$1.5 depending on the number of correct tickets	Base of \$0.15 plus bonus with a maximum of \$1 depending on the number of correct tickets
Number of participants used	503 participants	228 participants
Screening	Attention test and performance on a No-feedback market test case	No Screening

Table 24. Survey Data Collection Summary

4.1.2 Demographics

Demographic data was collected immediately after the consent. Table 25 gives a summary of the demographic responses. In both experiments there were slightly more male participants. While the majority of participants were under 30 years old, there were 42.5% over 30. The participants were generally well educated with 70.2% stating they had earned their college degree.

	Random Draws	Mirage Prone
What is your gender?		
Male	55.1%	53.7%
Female	44.9%	46.1%
What is your age?		
Under 18	Screened out	Screened out
18-29 years old	55.1%	57.1%
30-49 years old	37.2%	36.6%
50-64 years old	7.4%	5.0%
65 years and over	0.4%	0.9%
What is the highest level of education you have completed?		
some high school	1.8%	0.9%
high school graduate	7.6%	6.5%
some college/trade/technical/vocational training	21.3%	22.0%
college graduate	39.2%	39.1%
some postgraduate work	10.5%	8.1%
graduate degree	19.7%	23.0%
What was your major (if more than one then select the most recent)?		
Liberal Arts	11.1%	9.3%
Science	13.3%	14.3%
Mathematics/Statistics	3.0%	5.6%
Engineering	15.5%	17.1%
Business	10.1%	10.2%
Education	4.6%	3.7%
Other major	11.5%	10.2%
No Response	30.8%	29.6%

Table 25. Summary of Demographic Responses

4.1.3 Careless Responses

Some Turkers may use AMT as a means to provide substantial supplemental income.

Since workers are paid per task completed, there can be internally generated pressure on participants to complete as many tasks as possible in the time they have available to work.

Other researchers have reported careless responses in their experiments (Paolacci, Chandler, & Ipeirotis, 2010). This behavior was also inferred in the present study.

As mentioned, an attention test was included. This test, mixed in with the psychological factor Likert graded questions simply asked the participant to select 'Strongly Agree.'

17.1% of all participants failed to select 'strongly agree.' There was a significant Pearson Correlation of .206 (p-value of .002) between bonus amount and passing or failing the attention test.

Table 26 illustrates the impact of careless responses on a market. Participant number four receives a draw of D,D,E in an IP market. He proceeded to allocate all 100 tickets on state C.

	1st Draw	2nd Draw	3rd Draw	A	B	C	D	E	F
Participant 1	D	F	D	5	5	5	50	30	5
Participant 2	E	F	D	15	10	10	20	25	20
Participant 3	F	C	A	35	15	25	10	5	10
Participant 4	D	D	E	0	0	100	0	0	0
Participant 5	A	D	E	25	15	20	19	10	11
Participant 6	A	E	B	30	30	10	10	10	10
Participant 7	A	A	B	29	12	8	6	20	25
Participant 8	E	C	C	19	41	16	5	2	17
Participant 9	A	A	B	25	10	5	10	25	25
Participant 10	C	F	B	50	10	10	10	10	10
Participant 11	A	B	A	10	16	24	16	18	16
Participant 12	C	B	D	12	22	22	22	11	11
Participant 13	A	A	A	75	5	5	5	5	5
				330	191	260	183	171	165

Table 26. IP market with spurious input from Participant 5. The correct state is A

To check to determine if perhaps this participant simply got C and D mixed up, all of the responses from this participant are shown in Table 27. Notice a similar type of error in

their first IP market with 100 tickets allocated to state A while their draw was C,B,E. He also had similar illogical responses in the feedback markets. For example, in his final market he received a draw of D,F,F and the maximum public allocations were on states B and C. He, however, allocated 0 tickets on state F (the most likely from the draw) and 0 tickets on B (the most likely based on public information).

Draw Sets			Public Data						Allocations					
1st Draw	2nd Draw	3rd Draw	A	B	C	D	E	F	A	B	C	D	E	F
C	B	D	IP	IP	IP	IP	IP	IP	100	0	0	0	0	0
D	D	E	IP	IP	IP	IP	IP	IP	0	0	100	0	0	0
A	B	C	146	260	232	170	182	110	50	0	20	30	0	0
A	D	C	15	15	20	10	20	20	20	10	20	20	10	20
D	A	C	65	80	50	70	55	80	20	10	20	30	15	5
D	F	F	65	210	175	75	93	82	65	0	15	10	10	0

Table 27. All allocations from participant highlighted in Table 26.

4.1.4 Screening

Screening participants is an option to control careless responses. Two screening methods were employed in the random draws experiment. First, participants who failed the attention test (a Likert scale question which simply asked the participant to select 'Strongly Agree') were screened. Second, participants were screened based on their response in a benchmark no-feedback market. Screened participants completed the survey as would a non-screened participant. They were also paid the base compensation as well as the bonus. However, screened participant's data was not included in the general database and therefore was not incorporated into the markets.

In the random draws experiment, participants were presented a third no-feedback market. This market had a constant draw of ECE (that is, two E balls and one C ball). Theoretical distributions based on the four methods identified for no-feedback markets were calculated as shown in Table 28. The Wurtz distance between the participant's allocation and each of these distributions was then calculated. The minimum of these Wurtz distance was then compared to 0.25. If the minimum Wurtz distance was greater than 0.25 the participant's results were not stored in the main database and were effectively screened from later participants.

	A	B	C	D	E	F
Maximin	16	16	17	16	19	16
Pct	0	0	33	0	67	0
Bayes	8	8	19	8	49	8
EV	0	0	0	0	100	0

Table 28. Predicted responses to ECE draw for each allocation method.

4.2 Market Performance with Random Draws

This section will describe the results from the random draws experiment.

4.2.1 Information Collection in Markets with No-feedback

The ability of a market to collect information can be studied in the no-feedback markets.

As mentioned previously, market aggregation can be broken into two components:

collection of the individual analysis and judgment of private information; and a market level induction process to evaluate the collected information. The no-feedback markets can be used to study the first process since induction is not possible. Table 29 illustrates the comparison. The table shows nine draw sets and the allocations made by participants.

For example, the first participant received a draw of B,A,F and allocated tickets across states A-F as 32,32,1,2,1,32. The two step process outlined in section 3.1.1.5 can then be used on this distribution. Since the respondent allocated 32 tickets on states A, B and F the result is 'Not Significant.' Each of the nine respondents' allocations can be evaluated with the result of 1 'Correct,' 3 'Not Correct' and 5 'Not Significant.'

Correctness can be calculated for these responses as $1/3$, and decisiveness as $4/9$. This can be compared to the sum of all allocations shown at the bottom as 168, 188, 168, 114, 128, 134 for states A-F. This cumulative allocation can be characterized as 'Not Significant.' When multiple markets are considered then values for correctness and Decisiveness can be computed and compared to the correctness and decisiveness of the individuals.

Respondent	Draws			Allocations						Characterization
	1st	2nd	3rd	A	B	C	D	E	F	
1	B	A	F	32	32	1	2	1	32	Not Significant
2	A	B	A	36	16	12	12	12	12	Correct
3	C	F	E	10	10	15	15	25	25	Not Significant
4	C	A	D	20	20	20	20	10	10	Not Significant
5	E	C	D	20	20	20	20	10	10	Not Significant
6	A	A	A	10	10	20	20	20	20	Not Significant
7	C	A	B	20	10	50	5	5	10	Incorrect
8	B	B	C	10	60	10	10	5	5	Incorrect
9	E	E	C	10	10	20	10	40	10	Incorrect
	Total			168	188	168	114	128	134	Not Significant

Table 29. Illustration of individual versus market results.

Table 30 shows the decisiveness and correctness of the no-feedback market by stage (participant). In addition the performance of the individuals included in those market results is also shown. The individual results are calculated in the same method as the market results. The two stage inferential process described in Section 3.4.1 can be applied to the individual's allocation. These can then be used to calculate decisiveness and correctness values. The two decisiveness and correctness values can be compared with the Fisher's Exact Test and the resulting p-value is also shown.

Stage	Decisiveness			Correctness		
	Individuals	Markets	p-value	Individuals	Markets	p-value
1	33.3%	33.3%	0.5856	50.0%	50.0%	0.6378
2	37.5%	43.8%	0.2929	58.3%	57.1%	0.5738
3	38.2%	39.6%	0.4974	58.2%	57.9%	0.5949
4	37.0%	35.4%	0.4903	52.1%	70.6%	0.1341
5	35.8%	39.6%	0.3681	47.7%	63.2%	0.1666
6	35.4%	52.1%	0.0215	48.0%	68.0%	0.0577
7	34.2%	60.4%	0.0005	45.2%	65.5%	0.0399
8	35.4%	56.3%	0.0045	45.6%	77.8%	0.0019
9	36.6%	58.3%	0.0030	44.9%	75.0%	0.0028
10	36.0%	62.5%	0.0004	44.5%	73.3%	0.0030
11	36.6%	58.3%	0.0027	43.0%	71.4%	0.0043
12	35.9%	58.3%	0.0020	43.0%	71.4%	0.0041
13	36.5%	56.3%	0.0058	44.3%	81.5%	0.0002
14	36.0%	58.3%	0.0020	44.6%	82.1%	0.0001
15	36.5%	56.3%	0.0056	46.0%	92.6%	0.0000
16	37.0%	52.1%	0.0275	45.8%	100.0%	0.0000
17	36.9%	56.3%	0.0063	45.5%	100.0%	0.0000
18	37.3%	60.4%	0.0013	45.0%	100.0%	0.0000
19	36.8%	60.4%	0.0010	44.9%	100.0%	0.0000
20	36.6%	58.3%	0.0023	45.0%	100.0%	0.0000

Table 30. Summary of decisiveness and correctness of individual responses vs. no-feedback markets by market stage (e.g., number of participants) from the random draws experiment

As seen, after stage 6 the market outperforms the collection of individual responses in both decisiveness and correctness with statistical significance of .05.

4.2.2 Information Aggregation in Markets with Feedback

The information aggregation capability of markets can be studied by comparing the performance of markets with different types of feedback.

4.2.2.1 Modal Performance

First, the basic performance of the markets is examined. Table 31 shows the performance of the no-feedback and feedback markets described as correct, incorrect or not significant as described in section 3.1.1.5. Table 32 shows the same results expressed as

percentages. After 16 stages, no-feedback markets show no incorrect results. However, approximately 42% (20 out of 48) are unable to be judged as statistically significant.

Table 31 also shows the modal performance of the feedback markets. It is immediately noted that in contrast to the no-feedback markets, there are a large fraction of the results which are incorrect in the final stages of the markets. Nearly 18% of the markets end the test in an incorrect state. Also of note is that the percentage of markets judged as not significant is much smaller than in the no-feedback case (approximately 18% vs. 42% for the no-feedback markets).

Stage	No-Feedback			Feedback		
	Correct	Incorrect	Not Sig.	Correct	Incorrect	Not Sig.
1	8	8	32	15	17	57
2	12	9	27	19	15	55
3	11	8	29	28	15	46
4	12	5	31	29	13	47
5	12	7	29	31	20	38
6	17	8	23	37	20	32
7	19	10	19	41	18	30
8	21	6	21	43	19	27
9	21	7	20	42	14	33
10	22	8	18	43	15	31
11	20	8	20	45	16	28
12	20	8	20	45	17	27
13	22	5	21	47	16	26
14	23	5	20	48	16	25
15	25	2	21	51	16	22
16	25	0	23	53	14	22
17	27	0	21	55	15	19
18	29	0	19	56	16	17
19	29	0	19	56	17	16
20	28	0	20	56	16	17

Table 31. No-feedback and feedback market results as counts by market stage (e.g., number of participants) from the random draws experiment

Stage	No-Feedback				Feedback		
	Correct	Incorrect	Not Sig.		Correct	Incorrect	Not Sig.
1	16.7%	16.7%	66.7%		16.9%	19.1%	64.0%
2	25.0%	18.8%	56.3%		21.3%	16.9%	61.8%
3	22.9%	16.7%	60.4%		31.5%	16.9%	51.7%
4	25.0%	10.4%	64.6%		32.6%	14.6%	52.8%
5	25.0%	14.6%	60.4%		34.8%	22.5%	42.7%
6	35.4%	16.7%	47.9%		41.6%	22.5%	36.0%
7	39.6%	20.8%	39.6%		46.1%	20.2%	33.7%
8	43.8%	12.5%	43.8%		48.3%	21.3%	30.3%
9	43.8%	14.6%	41.7%		47.2%	15.7%	37.1%
10	45.8%	16.7%	37.5%		48.3%	16.9%	34.8%
11	41.7%	16.7%	41.7%		50.6%	18.0%	31.5%
12	41.7%	16.7%	41.7%		50.6%	19.1%	30.3%
13	45.8%	10.4%	43.8%		52.8%	18.0%	29.2%
14	47.9%	10.4%	41.7%		53.9%	18.0%	28.1%
15	52.1%	4.2%	43.8%		57.3%	18.0%	24.7%
16	52.1%	0.0%	47.9%		59.6%	15.7%	24.7%
17	56.3%	0.0%	43.8%		61.8%	16.9%	21.3%
18	60.4%	0.0%	39.6%		62.9%	18.0%	19.1%
19	60.4%	0.0%	39.6%		62.9%	19.1%	18.0%
20	58.3%	0.0%	41.7%		62.9%	18.0%	19.1%

Table 32. No-feedback and feedback market results as percentages by market stage (e.g., number of participants) from the random draws experiment.

The decisiveness and correctness results of the no-feedback and feedback markets are shown in Table 33. After just a few stages the decisiveness value of the feedback markets is descriptively larger than the no-feedback markets. After 13 stages the correctness of the no-feedback markets is larger than the correctness of the feedback markets. The values are compared as described in section 3.1.1.6 and the resulting p-value is shown in the table. After stage 15 the differences for both decisiveness and correctness are different at the 5% level. These results illustrate the fundamental tradeoff between these two market types. The feedback is enabling the induction

process which is allowing the markets to come to more conclusions about the correct state (decisiveness). However, the induction process, which makes the markets subject to information mirages, adversely impacts the correctness of those judgments.

Stage	Decisiveness			Correctness		
	No-FB	FB	p-value	No-FB	FB	p-value
1	33.3%	36.0%	0.4548	50.0%	46.9%	0.5400
2	43.8%	38.2%	0.3257	57.1%	55.9%	0.5756
3	39.6%	48.3%	0.2122	57.9%	65.1%	0.3951
4	35.4%	47.2%	0.1255	70.6%	69.0%	0.5840
5	39.6%	57.3%	0.0357	63.2%	60.8%	0.5415
6	52.1%	64.0%	0.1192	68.0%	64.9%	0.4969
7	60.4%	66.3%	0.3081	65.5%	69.5%	0.4433
8	56.3%	69.7%	0.0840	77.8%	69.4%	0.2931
9	58.3%	62.9%	0.3648	75.0%	75.0%	0.5995
10	62.5%	65.2%	0.4486	73.3%	74.1%	0.5640
11	58.3%	68.5%	0.1570	71.4%	73.8%	0.5043
12	58.3%	69.7%	0.1266	71.4%	72.6%	0.5503
13	56.3%	70.8%	0.0648	81.5%	74.6%	0.3382
14	58.3%	71.9%	0.0780	82.1%	75.0%	0.3212
15	56.3%	75.3%	0.0187	92.6%	76.1%	0.0550
16	52.1%	75.3%	0.0054	100.0%	79.1%	0.0090
17	56.3%	78.7%	0.0057	100.0%	78.6%	0.0053
18	60.4%	80.9%	0.0090	100.0%	77.8%	0.0029
19	60.4%	82.0%	0.0057	100.0%	76.7%	0.0021
20	58.3%	80.9%	0.0046	100.0%	77.8%	0.0034

Table 33. Decisiveness and correctness for feedback and no-feedback markets by market stage (e.g., number of participants) from the random draws experiment.

The feedback markets consist of two types of feedback: frequency and percentage. The modal results of these two are shown in Table 34 and Table 35.

Stage	Frequency				Percent		
	Correct	Incorrect	Not Sig.		Correct	Incorrect	Not Sig.
1	6	8	32		9	9	25
2	9	6	31		10	9	24
3	14	6	26		14	9	20
4	13	7	26		16	6	21
5	17	12	17		14	8	21
6	19	12	15		18	8	17
7	21	11	14		20	7	16
8	22	11	13		21	8	14
9	23	8	15		19	6	18
10	22	10	14		21	5	17
11	24	10	12		21	6	16
12	24	10	12		21	7	15
13	26	10	10		21	6	16
14	27	10	9		21	6	16
15	29	10	7		22	6	15
16	31	9	6		22	5	16
17	31	10	5		24	5	14
18	31	10	5		25	6	12
19	31	10	5		25	7	11
20	31	10	5		25	6	12

Table 34. Market results for frequency and percent markets as counts by market stage (e.g., number of participants) from the random draws experiment.

Stage	Frequency				Percent		
	Correct	Incorrect	Not Sig.		Correct	Incorrect	Not Sig.
1	13.0%	17.4%	69.6%		20.9%	20.9%	58.1%
2	19.6%	13.0%	67.4%		23.3%	20.9%	55.8%
3	30.4%	13.0%	56.5%		32.6%	20.9%	46.5%
4	28.3%	15.2%	56.5%		37.2%	14.0%	48.8%
5	37.0%	26.1%	37.0%		32.6%	18.6%	48.8%
6	41.3%	26.1%	32.6%		41.9%	18.6%	39.5%
7	45.7%	23.9%	30.4%		46.5%	16.3%	37.2%
8	47.8%	23.9%	28.3%		48.8%	18.6%	32.6%
9	50.0%	17.4%	32.6%		44.2%	14.0%	41.9%
10	47.8%	21.7%	30.4%		48.8%	11.6%	39.5%
11	52.2%	21.7%	26.1%		48.8%	14.0%	37.2%
12	52.2%	21.7%	26.1%		48.8%	16.3%	34.9%
13	56.5%	21.7%	21.7%		48.8%	14.0%	37.2%
14	58.7%	21.7%	19.6%		48.8%	14.0%	37.2%
15	63.0%	21.7%	15.2%		51.2%	14.0%	34.9%
16	67.4%	19.6%	13.0%		51.2%	11.6%	37.2%
17	67.4%	21.7%	10.9%		55.8%	11.6%	32.6%
18	67.4%	21.7%	10.9%		58.1%	14.0%	27.9%
19	67.4%	21.7%	10.9%		58.1%	16.3%	25.6%
20	67.4%	21.7%	10.9%		58.1%	14.0%	27.9%

Table 35. Market results for frequency and percent markets as percentage by market stage (e.g., number of participants) from the random draws experiment

Table 36 shows the decisiveness and correctness results for the two types of feedback markets. Descriptively, after 5 market stages the frequency feedback markets exhibit higher decisiveness. Descriptively, after 4 market stages the percent feedback markets exhibit higher correctness. Significance testing, as described in section 3.1.1.6 shows that the difference in correctness is significant at the 5% level for stages 15,16,17,18 and 20 and significant at the 10% level after stage 13. The difference in correctness is not statistically significant. At the small effect size observed (approximately 5% difference in correctness) it would require approximately 600 total feedback markets to observe statistical significance at the 10% level.

Stage	Decisiveness			Correctness		
	Freq	Pct	p-value	Freq	Pct	p-value
1	30.4%	41.9%	0.1837	42.9%	50.0%	0.4826
2	32.6%	44.2%	0.1828	60.0%	52.6%	0.4682
3	43.5%	53.5%	0.2322	70.0%	60.9%	0.3810
4	43.5%	51.2%	0.3040	65.0%	72.7%	0.4176
5	63.0%	51.2%	0.1794	58.6%	63.6%	0.4718
6	67.4%	60.5%	0.3229	61.3%	69.2%	0.3654
7	69.6%	62.8%	0.3259	65.6%	74.1%	0.3392
8	71.7%	67.4%	0.4166	66.7%	72.4%	0.4165
9	67.4%	58.1%	0.2472	74.2%	76.0%	0.5636
10	69.6%	60.5%	0.2490	68.8%	80.8%	0.2314
11	73.9%	62.8%	0.1839	70.6%	77.8%	0.3688
12	73.9%	65.1%	0.2510	70.6%	75.0%	0.4614
13	78.3%	62.8%	0.0851	72.2%	77.8%	0.4202
14	80.4%	62.8%	0.0529	73.0%	77.8%	0.4452
15	84.8%	65.1%	0.0280	74.4%	78.6%	0.4604
16	87.0%	62.8%	0.0079	77.5%	81.5%	0.4703
17	89.1%	67.4%	0.0121	75.6%	82.8%	0.3400
18	89.1%	72.1%	0.0374	75.6%	80.6%	0.4150
19	89.1%	74.4%	0.0625	75.6%	78.1%	0.5132
20	89.1%	72.1%	0.0374	75.6%	80.6%	0.4150

Table 36. Decisiveness and correctness for the two types of feedback markets studied - frequency and percent feedback by market stage (e.g., number of participants) from the random draws experiment

Figure 13 summarizes the descriptive results at the end of the experiment (after stage 20).

No-feedback markets offer a perfect correctness, but at the expense of limited decisiveness. Frequency feedback markets exhibit the greatest decisiveness, but at the cost of some errors. Percent feedback markets, which moderate the feedback to some extent, slightly improve the correctness at the cost of some decisiveness.

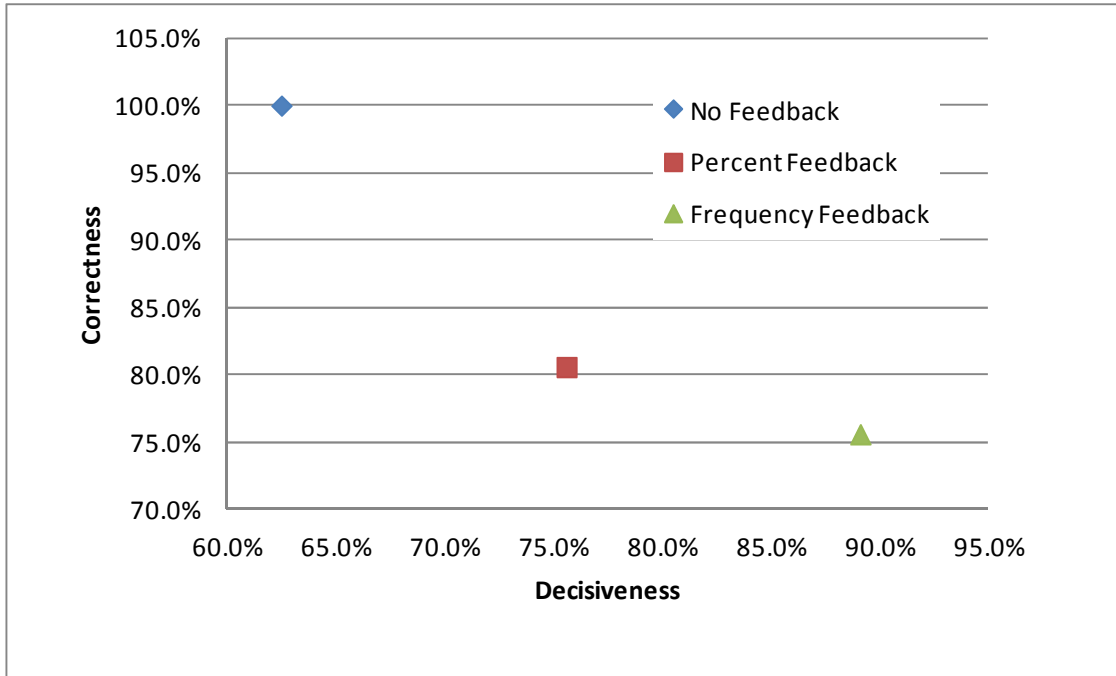


Figure 13. Summary of market performance illustrating the tradeoff between correctness and decisiveness after stage 20 for the random draws experiment.

4.2.2.2 Wurtz Distance to Final AIA

Table 37 shows the mean value of the Wurtz distance between the cumulative allocation at any given stage and the AIA distribution (as described in section 3.1.1.3). For each market, the AIA distribution is determined by calculating the Bayesian posterior probability of the cumulative ball draws at that stage. The Wurtz distance is then calculated between this distribution and the cumulative allocation. Finally, the mean value of this metric across all markets is calculated and is displayed in the table. The two distributions of metrics were compared using a Mann-Whitney-Wilcoxon non-parametric test and the p-value reported in the table. Descriptively, after stage 3, the feedback markets exhibit a lower Wurtz distance to the AIA than no-feedback markets. This difference is statistically significant at the 5% level after stage 14, and significant at

the 10% level after stage 11. This indicates that feedback markets are getting closer to the AIA distribution, which is one evaluation of the correct result.

Stage	No-Feedback	Feedback	p-value		Frequency	Percent	p-value
1	0.23576	0.23079	0.446		0.22463	0.23738	0.332
2	0.28738	0.29557	0.345		0.28688	0.30486	0.256
3	0.38687	0.37741	0.368		0.37014	0.38519	0.433
4	0.45423	0.43268	0.280		0.43087	0.43461	0.456
5	0.49727	0.48191	0.303		0.49130	0.47186	0.250
6	0.54540	0.52373	0.237		0.52547	0.52187	0.462
7	0.56816	0.55308	0.251		0.54846	0.55803	0.330
8	0.59958	0.59223	0.324		0.58619	0.59869	0.330
9	0.61472	0.60693	0.288		0.60328	0.61083	0.382
10	0.64252	0.62299	0.124		0.61519	0.63134	0.401
11	0.64898	0.62500	0.095		0.61731	0.63323	0.316
12	0.66247	0.63999	0.064		0.63204	0.64850	0.333
13	0.67711	0.65422	0.078		0.64611	0.66289	0.277
14	0.70328	0.67875	0.036		0.66650	0.69185	0.149
15	0.72647	0.70193	0.038		0.69303	0.71145	0.178
16	0.73792	0.71266	0.040		0.70376	0.72219	0.163
17	0.74668	0.71943	0.012		0.70957	0.72997	0.153
18	0.74960	0.71816	0.001		0.70496	0.73229	0.067
19	0.75679	0.72530	0.002		0.71123	0.74035	0.032
20	0.75972	0.72541	0.001		0.71174	0.74003	0.057

Table 37. Mean value of Wurtz distance between cumulative allocation and AIA distribution by market stage (e.g., number of participants) from the random draws experiment

Table 37 also shows the same information for the two types of feedback markets.

Descriptively, frequency feedback markets display a lower AIA Wurtz distance after stage 7. This difference is statistically significant at the 5% level at stage 19 and significant at the 10% level after stage 18.

4.2.2.3 Entropy

Entropy of the cumulative allocation was measured and is an indication of the certainty of the market. Table 38 gives the mean value of the entropy of the cumulative allocation across all markets. Feedback markets are seen to have lower entropy than no-feedback markets, with this difference being statistically significant at the 5% level (Mann-Whitney-Wilcoxon non-parametric test) after stage 3.

Stage	No FB	FB	p-value		Freq	Pct	p-value
1	2.1680	2.1640	0.4240		2.2088	2.1162	0.1457
2	2.3725	2.3213	0.3950		2.3253	2.3171	0.2517
3	2.4425	2.3474	0.0030		2.3516	2.3429	0.2287
4	2.4781	2.3938	0.0019		2.3973	2.3900	0.3155
5	2.4987	2.4212	0.0001		2.4198	2.4226	0.4494
6	2.5044	2.4263	0.0000		2.4235	2.4293	0.3791
7	2.5132	2.4430	0.0001		2.4432	2.4428	0.3916
8	2.5130	2.4523	0.0002		2.4485	2.4563	0.2491
9	2.5124	2.4604	0.0003		2.4591	2.4618	0.2925
10	2.5195	2.4650	0.0000		2.4591	2.4714	0.2189
11	2.5269	2.4682	0.0000		2.4632	2.4734	0.2623
12	2.5297	2.4740	0.0000		2.4687	2.4796	0.2623
13	2.5315	2.4759	0.0000		2.4678	2.4845	0.1757
14	2.5348	2.4789	0.0000		2.4713	2.4870	0.0994
15	2.5362	2.4787	0.0000		2.4682	2.4899	0.0542
16	2.5372	2.4826	0.0000		2.4721	2.4939	0.0450
17	2.5396	2.4870	0.0000		2.4775	2.4972	0.0490
18	2.5411	2.4869	0.0000		2.4756	2.4990	0.0315
19	2.5436	2.4881	0.0000		2.4762	2.5007	0.0242
20	2.5454	2.4875	0.0000		2.4740	2.5020	0.0109

Table 38. Mean entropy by stage by market stage (e.g., number of participants) from the random draws experiment

Table 38 also gives the mean entropy of the cumulative allocation for the feedback markets broken down by feedback type. Descriptively, after stage 8 the frequency

markets exhibit lower entropy. This difference is statistically significant at the 5% level after stage 16 and significant at the 10% level after stage 14.

The entropy distributions are also of interest. Table 39 shows the means as well as the standard deviations of the distributions for the feedback and no-feedback markets by market stage. In addition to the mean differences just described, the table shows that the standard deviation of the feedback markets is approximately twice as large as for the no-feedback markets. The Levene's test (Levene, 1960) p-value is also shown which test for the homogeneity of variance of the two distributions. This test is significant at the 5% level after stage 2, indicating a non-homogenous variance.

Stage	No-Feedback		Feedback		Levene's
	Mean	SD	Mean	SD	p-value
1	2.1680	0.5188	2.1640	0.5257	0.9133
2	2.3725	0.1799	2.3213	0.3090	0.0162
3	2.4425	0.1524	2.3474	0.2691	0.0262
4	2.4781	0.1060	2.3938	0.2015	0.0036
5	2.4987	0.0700	2.4212	0.1623	0.0007
6	2.5044	0.0665	2.4263	0.1448	0.0005
7	2.5132	0.0525	2.4430	0.1298	0.0001
8	2.5130	0.0474	2.4523	0.1166	0.0002
9	2.5124	0.0617	2.4604	0.1047	0.0020
10	2.5195	0.0531	2.4650	0.0918	0.0011
11	2.5269	0.0427	2.4682	0.0927	0.0002
12	2.5297	0.0403	2.4740	0.0852	0.0000
13	2.5315	0.0362	2.4759	0.0807	0.0003
14	2.5348	0.0358	2.4789	0.0774	0.0001
15	2.5362	0.0347	2.4787	0.0733	0.0001
16	2.5372	0.0376	2.4826	0.0668	0.0003
17	2.5396	0.0387	2.4870	0.0623	0.0002
18	2.5411	0.0406	2.4869	0.0636	0.0002
19	2.5436	0.0372	2.4881	0.0652	0.0002
20	2.5454	0.0350	2.4875	0.0651	0.0002

Table 39. Characteristics of the entropy distribution for no-feedback and feedback markets by market stage (e.g., number of participants) from the random draws experiment

Table 40 shows the mean and standard deviation of the entropy distributions of the two types of feedback markets over all stages. No difference is noticed in the standard deviations and the Levene's test p-value is not significant at the 10% level.

Stage	Frequency		Percent		Levene's
	Mean	SD	Mean	SD	p-value
1	2.2088	0.4742	2.1162	0.5776	0.402175
2	2.3253	0.3156	2.3171	0.3054	0.691826
3	2.3516	0.2710	2.3429	0.2703	0.738352
4	2.3973	0.2093	2.3900	0.1952	0.927856
5	2.4198	0.1815	2.4226	0.1410	0.831904
6	2.4235	0.1583	2.4293	0.1307	0.985383
7	2.4432	0.1351	2.4428	0.1256	0.638948
8	2.4485	0.1244	2.4563	0.1089	0.769353
9	2.4591	0.1062	2.4618	0.1042	0.388213
10	2.4591	0.0984	2.4714	0.0849	0.933079
11	2.4632	0.1009	2.4734	0.0839	0.99216
12	2.4687	0.0924	2.4796	0.0774	0.819908
13	2.4678	0.0894	2.4845	0.0703	0.723931
14	2.4713	0.0818	2.4870	0.0725	0.896783
15	2.4682	0.0780	2.4899	0.0671	0.910567
16	2.4721	0.0700	2.4939	0.0621	0.996285
17	2.4775	0.0636	2.4972	0.0598	0.983483
18	2.4756	0.0668	2.4990	0.0584	0.624467
19	2.4762	0.0670	2.5007	0.0615	0.562066
20	2.4740	0.0641	2.5020	0.0637	0.663406

Table 40. Characteristics of the entropy distribution for frequency feedback and percent feedback markets by market stage (e.g., number of participants) from the random draws experiment.

Figure 14, a histogram of the feedback and no-feedback entropy distributions after stage 20, allows a visual comparison. For the no-feedback markets all but 1 of the final entropy values lie in the range of 2.5 to 2.58 (with the lowest final entropy for a no-feedback market in the 2.36 bin). By contrast, the feedback markets exhibit a large tail of low entropy values ranging down to a low in the 2.22 bin. This illustrates the induction process as the market agrees on a correct value and lowers the entropy.

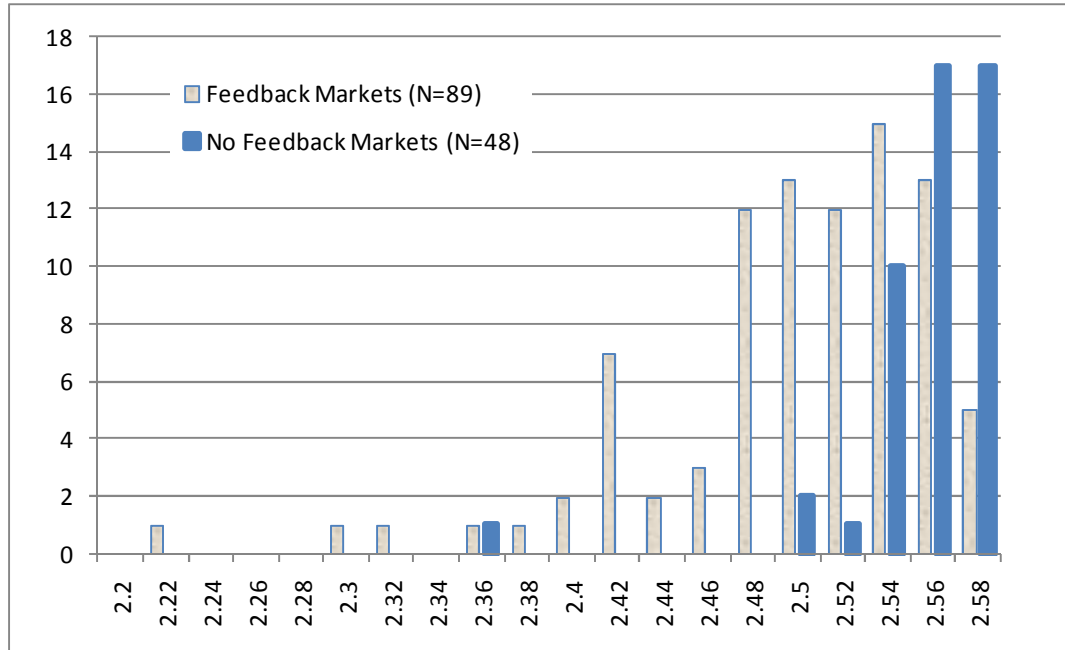


Figure 14. Histogram of final entropy distributions (after stage 20) for the random draws experiment.

4.2.2.4 Relationship between Mean Entropy and Decisiveness

An inversely monotonic relationship is expected between mean entropy and decisiveness.

Figure 15 shows the relationship as measured after stage 20.

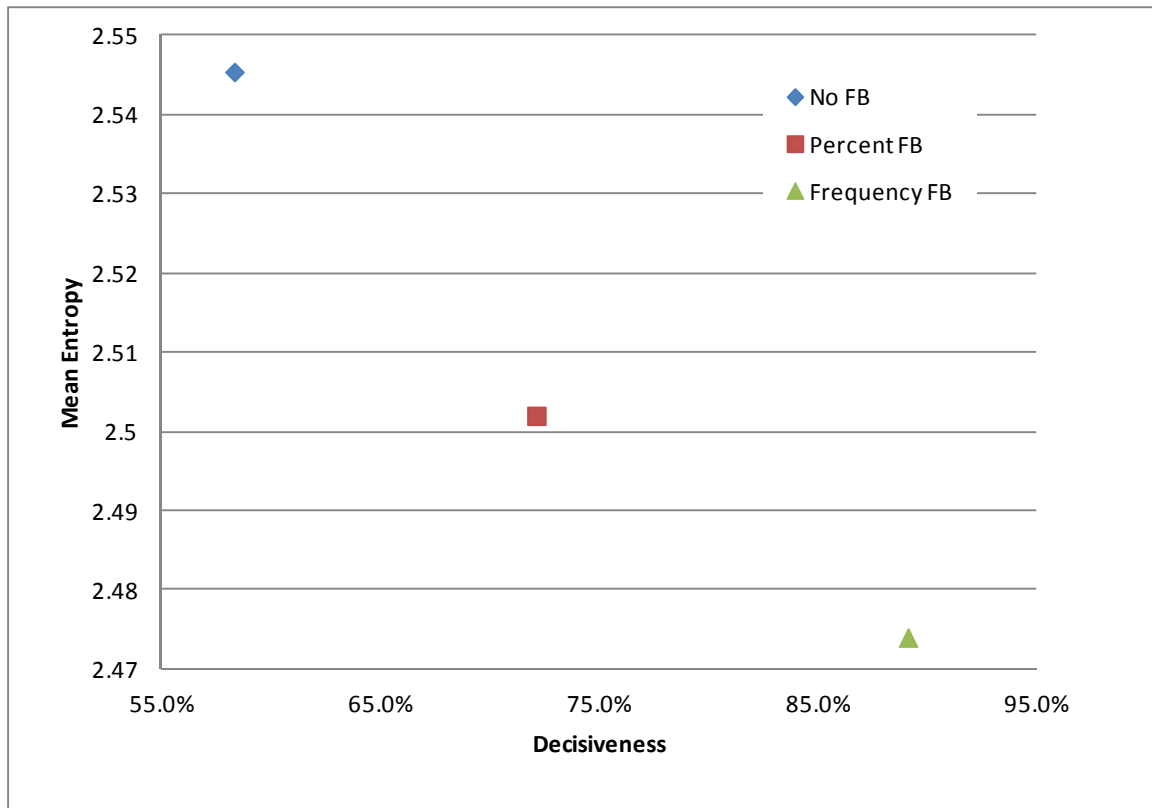


Figure 15. Relationship between mean entropy and decisiveness for each market type (after stage 20) for the random draws experiment.

4.2.3 Mirage Tendencies in Feedback Markets

The tendency of the induction process in feedback to create information mirages is studied.

4.2.3.1 Qualitative Analysis of Incorrect Feedback Markets

Examining individual markets can illustrate some of the dynamics taking place in markets which come to a statistically significant but incorrect conclusion (see section 3.1.1.5). Table 41 gives the relevant details for a frequency feedback market. The correct state for this market is E. The histogram in Table 42 shows the correct state has a total of 17 draws while state D, the eventual erroneous prediction, has 10 draws. The mirage begins early with the second participant (P2). He receives two draws of state D

and makes a rational allocation of 50 tickets on that state. Participant 3, who receives a draw of BFB places 30 tickets on D even though it was not in his draw as a result of the large public allocation on that state. Participant 4, who receives a strong draw for the correct state, makes a temperate allocation to that state and puts 17 tickets on state D even though it wasn't in his draw. Participant 13 is also significant in that he allocates 70 tickets on state D given a single draw. An important characteristic of this draw set is the lack of draws on the correct state from participant 3 through participant 10. This allows the mirage to build as 8 draws to the error state occur in that span.

Frequency Market Bubble (DG=11)										
Respondent	Draws			Allocations						
	1st	2nd	3rd	A	B	C	D	E	F	
1	F	C	C	3	10	25	3	4	55	
2	D	D	F	10	10	10	50	10	10	
3	B	F	B	12	15	8	30	5	30	
4	B	E	E	5	9	19	17	35	15	
5	E	B	A	30	20	10	10	20	10	
6	F	E	C	5	13	20	35	15	12	
7	C	C	E	15	15	15	20	15	20	
8	E	C	E	10	10	10	30	10	30	
9	A	A	B	23	12	13	20	12	20	
10	E	C	F	14	14	17	15	17	23	
11	D	D	E	0	0	0	50	50	0	
12	E	C	B	10	30	20	10	20	10	
13	B	C	E	3	3	3	70	8	13	
14	A	B	F	15	15	15	20	15	20	
15	D	F	F	0	0	0	34	0	66	
16	E	E	D	5	5	5	10	50	25	
17	A	C	D	20	17	21	18	12	12	
18	D	D	C	22	20	18	12	13	15	
19	E	E	A	20	10	10	10	40	10	
20	E	D	E	0	0	0	40	60	0	
				222	228	239	504	411	396	

Table 41. Frequency market which exhibits an information mirage.

Histogram	
<i>State</i>	<i>Frequency</i>
A	6
B	8
C	11
D	10
E	17
F	8

Table 42. Histogram of states for draw set shown in Table 41.

Table 43 illustrates a mirage in a percent market. The correct state is D and the erroneous mode occurs on state C. The mirage begins with the first participant who receives a strong signal for state C with two draws. He makes an allocation which closely resembles the Bayesian posterior probability (since there is no-feedback at this point the probabilities can be calculated as in a no-feedback market and are 49% for state C, 19.6% for state F and 7.8% for the other states). This initial allocation of 50 tickets on state C pull 92 tickets from participants 2-5 even though none of them received a draw of C. Participant 4 is significant in that he ignored his strong private information of two draws for state E, the correct state, and placed the largest number of tickets on states B and C. Participant 8 is also significant in that he received another strong signal for state E but also placed the largest number of tickets on other states.

Percent Market Bubble (DG=11)									
Respondent	Draws			Allocations					
	1st	2nd	3rd	A	B	C	D	E	F
1	F	C	C	5	5	50	5	5	30
2	D	D	F	20	30	20	20	10	0
3	B	F	B	6	50	25	6	3	10
4	B	E	E	10	30	30	10	10	10
5	E	B	A	17	17	17	15	17	17
6	F	E	C	15	10	15	10	15	35
7	C	C	E	15	20	30	10	15	10
8	E	C	E	0	20	30	0	20	30
9	A	A	B	10	30	35	0	5	20
10	E	C	F	20	10	15	30	15	10
11	D	D	E	5	5	10	50	20	10
12	E	C	B	10	10	40	10	10	20
13	B	C	E	11	19	26	13	12	19
14	A	B	F	15	20	30	10	10	15
15	D	F	F	0	20	60	0	0	20
16	E	E	D	3	4	3	30	30	30
17	A	C	D	25	0	30	45	0	0
18	D	D	C	10	10	30	30	10	10
19	E	E	A	15	15	25	15	15	15
20	E	D	E	10	35	25	15	5	10
				222	360	546	324	227	321

Table 43. Percentage feedback market which exhibits an information mirage.

Histogram	
State	Frequency
A	6
B	8
C	11
D	10
E	17
F	8

Table 44. Histogram of the states in the draw set in Table 43.

4.2.3.2 Market Results as a Function of Draw Set Characteristics

In section 3.4.1 a metric the 'Sum of AIA distances' was introduced to characterize draw sets. A larger sum of AIA distances indicates a draw set which is more prone to information mirages because the random nature of the draws takes longer to achieve an AIA distribution which gives high probability to the correct state. This metric was

compared to the Wurtz distance between final AIA and the cumulative allocation which is a measure of how well the market is performing (with perfect information assimilation the cumulative allocation will equal AIA). Figure 16 shows a scatter plot of these two metrics for each market in the random draws experiment broken out as feedback (circles) vs. no-feedback markets (asterisks). The dotted blue (feedback) and solid red (no-feedback) lines show the result of linear regression fit to the data points. For feedback markets the slope coefficient is statistically significant at the 5% level ($p=0.0112$). For the no-feedback markets the slope coefficient is not statistically significant at the 5% level ($p=0.8666$). This indicates the hypothesized relationship exists for feedback markets and, as expected, does not exist for no-feedback markets.

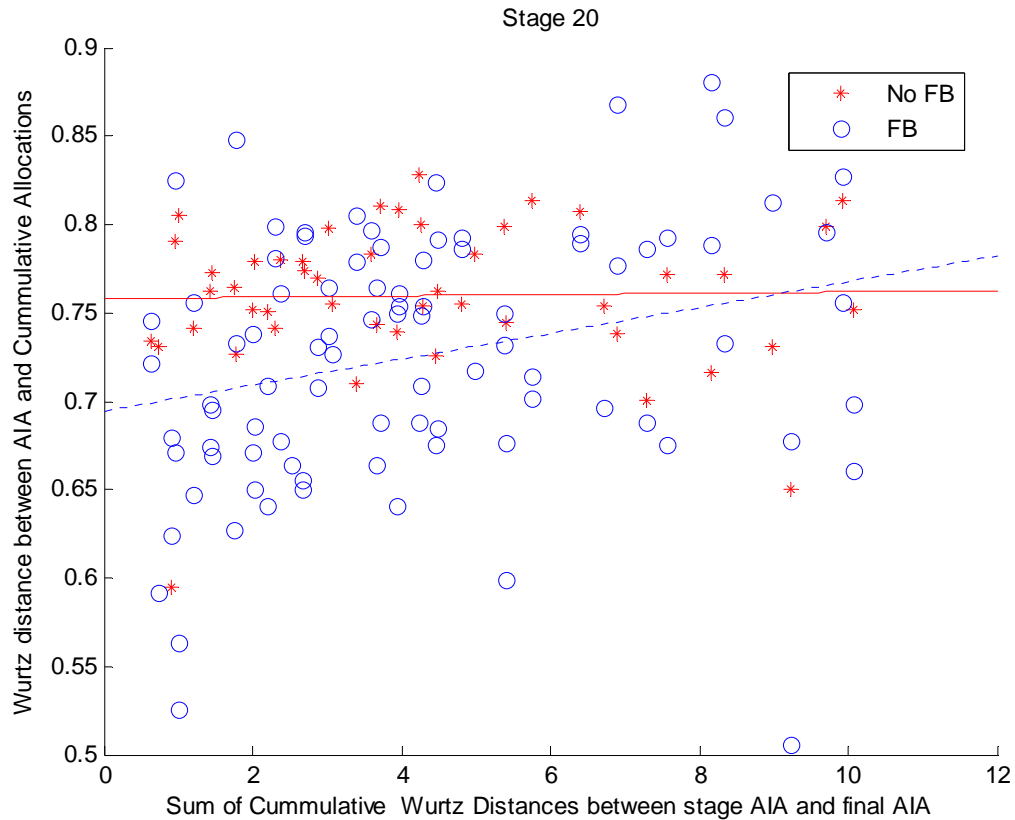


Figure 16. Wurtz distance between AIA and cumulative allocation as a function of the Sum of AIA distances after stage 20 for all markets in the random draws experiment.

4.3 Market Performance under Mirage Prone Conditions

The results of the markets under the mirage prone draw set condition is examined. A total of 91 markets were completed in the experiment. Each market had 13 stages. The breakdown was 33 no-feedback markets and 58 feedback markets (30 frequency feedback and 28 percent feedback).

4.3.1 Market's Performance with Mirage Prone Draw Sets

The market's performance with the mirage prone draw set is examined in the next sections

4.3.1.1 Modal Performance

The performance of the mode is examined first. Table 45 gives modal performance of the no-feedback and feedback markets expressed as counts. Table 46 shows the same information expressed as percentages. The allocations are judged as correct, incorrect or not significant as explained in section 3.1.1.5. Both markets had significant improvements in the thirteenth stage due to the draw of three balls on the correct state (see draw table in section 3.2.3).

Stage	No-Feedback				Feedback		
	Correct	Incorrect	Not Sig		Correct	Incorrect	Not Sig
1	1	20	12		0	26	32
2	1	14	18		0	25	33
3	0	14	19		0	21	37
4	0	19	14		0	29	29
5	0	22	11		0	30	28
6	2	19	12		0	25	33
7	3	17	13		2	27	29
8	3	16	14		1	27	30
9	6	12	15		2	26	30
10	5	12	16		0	20	38
11	6	8	19		3	20	35
12	5	10	18		3	22	33
13	13	7	13		8	20	30

Table 45. Performance of no-feedback and feedback markets under the mirage prone draw set expressed as counts by market stage (e.g., number of participants).

Stage	No-Feedback				Feedback		
	Correct	Incorrect	Not Sig		Correct	Incorrect	Not Sig
1	3.0%	60.6%	36.4%		0.0%	44.8%	55.2%
2	3.0%	42.4%	54.5%		0.0%	43.1%	56.9%
3	0.0%	42.4%	57.6%		0.0%	36.2%	63.8%
4	0.0%	57.6%	42.4%		0.0%	50.0%	50.0%
5	0.0%	66.7%	33.3%		0.0%	51.7%	48.3%
6	6.1%	57.6%	36.4%		0.0%	43.1%	56.9%
7	9.1%	51.5%	39.4%		3.4%	46.6%	50.0%
8	9.1%	48.5%	42.4%		1.7%	46.6%	51.7%
9	18.2%	36.4%	45.5%		3.4%	44.8%	51.7%
10	15.2%	36.4%	48.5%		0.0%	34.5%	65.5%
11	18.2%	24.2%	57.6%		5.2%	34.5%	60.3%
12	15.2%	30.3%	54.5%		5.2%	37.9%	56.9%
13	39.4%	21.2%	39.4%		13.8%	34.5%	51.7%

Table 46. Performance of no-feedback and feedback markets under the mirage prone draw set expressed as percentages by market stage (e.g., number of participants)..

Table 47 gives the decisiveness and correctness values as calculated from these basic results. Descriptively, the no-feedback markets have higher decisiveness than feedback markets although this result is not significant at the 5% or 10% levels. No-feedback markets have higher correctness than feedback markets with the result being significant at the 5% level after stages 9,10,11 and 13.

Stage	Decisiveness				Correctness		
	No-Feedback	Feedback	p-value		No-Feedback	Feedback	p-value
1	63.6%	44.8%	0.065		4.8%	0.0%	0.447
2	45.5%	43.1%	0.500		6.7%	0.0%	0.375
3	42.4%	36.2%	0.357		0.0%	0.0%	1.000
4	57.6%	50.0%	0.317		0.0%	0.0%	1.000
5	66.7%	51.7%	0.122		0.0%	0.0%	1.000
6	63.6%	43.1%	0.048		9.5%	0.0%	0.203
7	60.6%	50.0%	0.225		15.0%	6.9%	0.325
8	57.6%	48.3%	0.263		15.8%	3.6%	0.174
9	54.5%	48.3%	0.361		33.3%	7.1%	0.030
10	51.5%	34.5%	0.086		29.4%	0.0%	0.014
11	42.4%	39.7%	0.484		42.9%	13.0%	0.050
12	45.5%	43.1%	0.500		33.3%	12.0%	0.112
13	60.6%	48.3%	0.180		65.0%	28.6%	0.013

Table 47. Decisiveness and correctness for the feedback and no-feedback markets by stage for the mirage prone draw set by market stage (e.g., number of participants)..

Table 48 compares the decisiveness and correctness of the no-feedback and feedback markets after 13 stages with the mirage prone draw set. Descriptively, the decisiveness of no-feedback markets has increased slightly while the correctness has decreased somewhat. By contrast, the feedback markets show dramatic reductions in both decisiveness and correctness. The values are tested for statistical significance with the Fisher's exact test as described in section 3.4.3. As indicated in the table, for no-feedback markets the values of decisiveness and correctness are not different at the 5% level while both the correctness and decisiveness are different at the 5% level for feedback markets.

		Random Draws	Mirage Prone	p-value
Decisiveness	No-feedback	56.3%	60.6%	.436
	Feedback	70.8%	48.3%	.005
Correctness	No-feedback	81.5%	65.0%	.173
	Feedback	74.6%	28.6%	<.001

Table 48. Comparison of decisiveness and correctness after 13 stages for feedback and no-feedback markets in the mirage prone draw set experiment.

4.3.1.2 Impact of Feedback Type

Table 49 gives the modal performance of the frequency and percentage markets under the mirage prone draw set expressed as counts. Table 50 gives same information expressed as percentages.

Stage	Frequency				Percent		
	Correct	Incorrect	Not Sig		Correct	Incorrect	Not Sig
1	0	15	15		0	11	17
2	0	13	17		0	12	16
3	0	10	20		0	11	17
4	0	14	16		0	15	13
5	0	18	12		0	12	16
6	0	15	15		0	10	18
7	1	15	14		1	12	15
8	0	14	16		1	13	14
9	0	15	15		2	11	15
10	0	10	20		0	10	18
11	2	10	18		1	10	17
12	2	10	18		1	12	15
13	5	8	17		3	12	13

Table 49. Performance of frequency and percent feedback markets with mirage prone draw set expressed as counts by market stage (e.g., number of participants).

	No-Feedback				Feedback		
Stage	Correct	Incorrect	Not Sig		Correct	Incorrect	Not Sig
1	0.0%	50.0%	50.0%		0.0%	39.3%	60.7%
2	0.0%	43.3%	56.7%		0.0%	42.9%	57.1%
3	0.0%	33.3%	66.7%		0.0%	39.3%	60.7%
4	0.0%	46.7%	53.3%		0.0%	53.6%	46.4%
5	0.0%	60.0%	40.0%		0.0%	42.9%	57.1%
6	0.0%	50.0%	50.0%		0.0%	35.7%	64.3%
7	3.3%	50.0%	46.7%		3.6%	42.9%	53.6%
8	0.0%	46.7%	53.3%		3.6%	46.4%	50.0%
9	0.0%	50.0%	50.0%		7.1%	39.3%	53.6%
10	0.0%	33.3%	66.7%		0.0%	35.7%	64.3%
11	6.7%	33.3%	60.0%		3.6%	35.7%	60.7%
12	6.7%	33.3%	60.0%		3.6%	42.9%	53.6%
13	16.7%	26.7%	56.7%		10.7%	42.9%	46.4%

Table 50. Performance of frequency and percent feedback markets with mirage prone draw set expressed as percentages by market stage (e.g., number of participants).

Table 51 shows the values of decisiveness and correctness as calculated from the results shown in Table 49 and Table 50. Descriptively, the percentage markets have slightly higher decisiveness and frequency markets have slightly higher correctness. Both of these results are opposite to the relationship seen in the random draws experiment. However, both of these results are not statistically significant.

Stage	Decisiveness				Correctness		
	Frequency	Percent	p-value		Frequency	Percent	p-value
1	50.0%	39.3%	0.290		0.0%	0.0%	1.000
2	43.3%	42.9%	0.590		0.0%	0.0%	1.000
3	33.3%	39.3%	0.421		0.0%	0.0%	1.000
4	46.7%	53.6%	0.397		0.0%	0.0%	1.000
5	60.0%	42.9%	0.149		0.0%	0.0%	1.000
6	50.0%	35.7%	0.203		0.0%	0.0%	1.000
7	53.3%	46.4%	0.397		6.3%	7.7%	0.704
8	46.7%	50.0%	0.504		0.0%	7.1%	0.500
9	50.0%	46.4%	0.496		0.0%	15.4%	0.206
10	33.3%	35.7%	0.534		0.0%	0.0%	1.000
11	40.0%	39.3%	0.584		16.7%	9.1%	0.534
12	40.0%	46.4%	0.410		16.7%	7.7%	0.469
13	43.3%	53.6%	0.303		38.5%	20.0%	0.255

Table 51. Comparison of decisiveness and correctness by stage for the two types of feedback under the mirage prone draw set by market stage (e.g., number of participants)..

4.3.2 Evaluation for Information Mirages

The data characteristics which could identify information mirages is examined.

4.3.2.1 AIA Wurtz Distance

Table 52 gives the mean of the Wurtz distance between the cumulative allocation and the AIA distribution across the various stages. The distribution of Wurtz distance values are compared for a difference in the mean with the Mann-Whitney-Wilcoxon non-parametric test and the p-value reported in the table. In contrast to the result in the random draws experiment (Table 37) the no-feedback markets actually produce a lower Wurtz distance. This result is statistically significant at the 5% level after stage 11. This indicates that no-feedback markets are producing a more accurate result.

Mean Wurtz distance between AIA and Cumulative allocation			
Stage	No-Feedback	Feedback	p-value
1	0.2215	0.2822	0.031
2	0.3319	0.3480	0.280
3	0.3651	0.3490	0.251
4	0.5138	0.5095	0.401
5	0.6071	0.5886	0.410
6	0.5375	0.5203	0.318
7	0.4530	0.4442	0.334
8	0.3851	0.3777	0.404
9	0.4723	0.4842	0.147
10	0.4291	0.4384	0.155
11	0.6823	0.7037	0.006
12	0.5867	0.6036	0.024
13	0.7677	0.7945	0.004

Table 52. Mean Wurtz distance between the AIA distribution and the final cumulative ticket allocation for feedback and no-feedback markets by stage (No-feedback N=33; Feedback N=58)

Figure 17 shows a histogram of these distributions after stage 13 and may be able to give some insight into the differences. Notice that the smallest as well as largest Wurtz distances come from feedback markets. One explanation is that for the very low Wurtz distances, the induction process is driving the market result towards the actual AIA distribution. However, this induction can create an information mirage which would drive the market towards a large Wurtz distance (away from the correct AIA distribution).

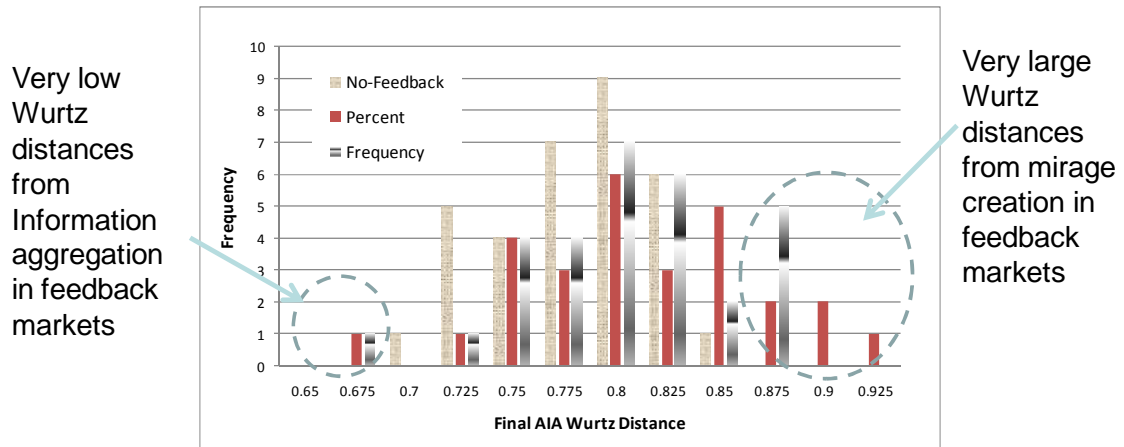


Figure 17. Wurtz distance between cumulative allocation at stage 13 and the AIA distribution for no-feedback and feedback markets in the mirage prone draw set experiment.

4.3.2.2 Variance in Entropy Distributions

Variance in the entropy distributions can provide additional insight into the information mirages. Table 53 shows the mean and standard deviations of the entropy distributions for the no-feedback and feedback markets for each stage. The feedback markets have a lower mean and larger standard deviation. The differences in the means is statistically significant at the 5% level after stage 9. Levene's test is used to judge the difference in the variance of the distributions. After stage 10 the two distributions have statistically significant difference in their variance at the 5% level. The markets with feedback have greater variance than the no-feedback markets.

Stage	Mean			Standard Deviation		
	No FB	FB	MW p-value	No FB	FB	Levene p-value
1	2.0776	2.1238	0.0293	0.3995	0.5963	0.2373
2	2.3377	2.2872	0.2152	0.1857	0.4292	0.0811
3	2.4185	2.4015	0.2972	0.1376	0.2300	0.2370
4	2.4205	2.3946	0.1964	0.1066	0.2413	0.0641
5	2.4574	2.4178	0.4458	0.0873	0.2025	0.0641
6	2.4886	2.4505	0.3324	0.0639	0.1701	0.0756
7	2.5045	2.4674	0.0726	0.0550	0.1253	0.1113
8	2.5162	2.4859	0.0968	0.0443	0.1090	0.1326
9	2.5296	2.5003	0.0213	0.0316	0.0838	0.0809
10	2.5404	2.5092	0.0036	0.0258	0.0726	0.0275
11	2.5428	2.5144	0.0045	0.0227	0.0614	0.0094
12	2.5482	2.5193	0.0006	0.0196	0.0511	0.0034
13	2.5413	2.5191	0.0032	0.0224	0.0479	0.0117

Table 53. Mean entropy by stage (e.g., number of participants) for feedback and no-feedback markets under the mirage prone draw set (No-feedback N=33; Feedback N=58)

Figure 18 is a box plot of the two distributions by stage illustrating the differences. This shows that in addition to wider standard deviations, the large outliers primarily are feedback markets.

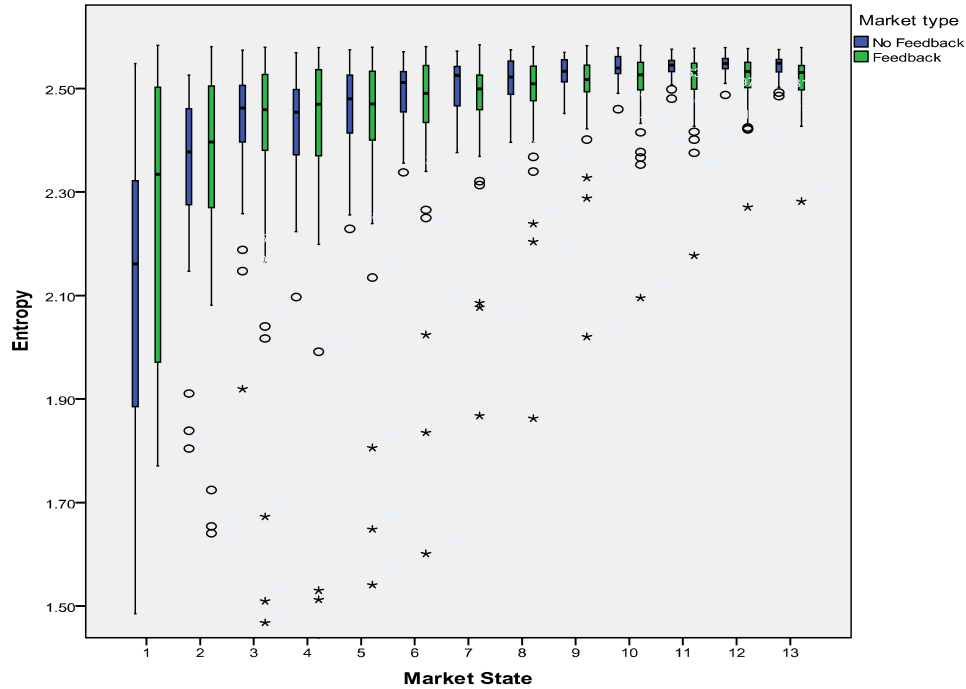


Figure 18. Entropy distributions for no-feedback and feedback markets by market stage (e.g., number of participants). Key: The dark line in the middle of the boxes is the median. The top and bottom of the box indicates the 75th and 25th percentiles, respectively. The T-bars, or whiskers, extend to 1.5 times the height of the box (if the data are distributed normally, approximately 95% or the data are expected to lie between the inner fences). The circles are outliers beyond the fences. Asterisks or stars are extreme outliers that have values more than three times the height of the boxes.

The inductive process present in feedback markets could lead to this type of effect. As a feedback market's inductive processes favors a particular state - whether it is a correct state or an information mirage - this would drive the entropy lower. The mirage prone draw set has elements that could drive this inductive process both in the direction of the mirage or in the direction of a correct answer. This conflict could result in the greater variance of entropy in feedback markets observed.

4.3.2.3 Relationship of Cumulative Allocation to Draw Set Characteristics

The relationship between the average allocation of tickets and the draw characteristics can shed some light on the information mirage effect. Table 54 shows the mirage prone

draw set used and the sum of draws by state per stage. The A state, the urn correct value, has the largest sum (12) with the D state having a total of 7 draws.

Draws			Sum of Draws						
1st	2nd	3rd	A	B	C	D	E	F	
D	F	D	0	0	0	2	0	1	
E	F	D	0	0	0	3	1	2	
F	C	A	1	0	1	3	1	3	
D	D	E	1	0	1	5	2	3	
A	D	E	2	0	1	6	3	3	
A	E	B	3	1	1	6	4	3	
A	A	B	5	2	1	6	4	3	
E	C	C	5	2	3	6	5	3	
A	A	B	7	3	3	6	5	3	
C	F	B	7	4	4	6	5	4	
A	B	A	9	5	4	6	5	4	
C	B	D	9	6	5	7	5	4	
A	A	A	12	6	5	7	5	4	

Table 54. Mirage prone draw set and the sum of draws by stage

Table 55 shows the sum of order weighted draws. In an information mirage, earlier draws are more influential since allocations which are a result of those draws can cause later participants to allocate more tickets to that state. The weighting factor studied here is a simple linear function as shown, with the first draw weighted at 13 (the total number of stages) and the final draw weighted at 1.

Draws					Sum of Weighted Draws					
1st	2nd	3rd		Weighting	A	B	C	D	E	F
D	F	D		13	0	0	0	26	0	13
E	F	D		12	0	0	0	38	12	25
F	C	A		11	11	0	11	38	12	36
D	D	E		10	11	0	11	58	22	36
A	D	E		9	20	0	11	67	31	36
A	E	B		8	28	8	11	67	39	36
A	A	B		7	42	15	11	67	39	36
E	C	C		6	42	15	23	67	45	36
A	A	B		5	52	20	23	67	45	36
C	F	B		4	52	24	27	67	45	40
A	B	A		3	58	27	27	67	45	40
C	B	D		2	58	29	29	69	45	40
A	A	A		1	61	29	29	69	45	40

Table 55. Mirage prone draw set and the sum of order weighted draws by stage

The average cumulative allocation of all participants passing the attention test are then calculated by stage and by market type as shown in Table 56 and Table 57. For example, after 13 stages, participants had allocated an average of 278.3 tickets on state A in the no-feedback markets. From these tables it is seen that in no-feedback markets, the largest cumulative allocation after the final stage is on state A, the urn correct state and the state with the largest sum of draws (12 total draws as shown in Table 54). By contrast the largest average cumulative allocation after stage 13 in the feedback markets was 267.4 for state D. State D has the largest order weighted sum of draws (69) as shown in Table 55.

A	B	C	D	E	F
10.6	8.7	10.5	39.9	9.9	20.3
23.4	20.1	28.6	58.8	31.1	37.9
41.5	31.1	55.8	71.9	44.2	55.5
52.0	43.0	70.5	108.2	61.9	64.4
72.3	57.3	88.0	124.6	78.6	79.2
97.9	75.2	101.2	137.7	97.7	89.9
128.2	96.4	114.3	148.8	111.1	100.5
141.4	108.7	140.3	165.0	131.7	111.0
171.5	128.0	151.5	179.1	145.2	122.8
183.8	148.7	174.8	192.5	156.4	141.9
217.6	166.6	189.5	205.6	166.8	151.9
230.7	184.3	217.0	223.8	178.8	163.5
278.3	195.0	227.1	235.2	188.2	174.3

Table 56. Mean cumulative allocations by stage (e.g., number of participants) for no-feedback markets (N=33)

A	B	C	D	E	F
10.6	11.7	12.4	34.8	13.0	17.5
20.8	23.0	24.8	59.3	32.7	39.4
38.6	36.2	43.1	74.5	46.6	61.1
49.5	47.9	55.5	109.2	62.1	75.8
66.7	59.0	67.2	134.5	83.2	89.4
87.5	76.4	80.2	150.5	102.4	103.1
115.5	96.3	89.4	167.1	116.9	114.7
128.1	110.9	113.4	181.9	135.4	130.3
153.8	128.7	126.7	197.7	148.6	144.3
166.3	145.3	145.4	215.7	161.8	164.8
194.4	163.9	158.1	230.8	174.5	177.7
209.7	182.3	178.4	251.5	187.4	189.8
243.7	193.4	190.7	267.4	202.1	201.7

Table 57. Mean cumulative allocations by stage (e.g., number of participants) for feedback markets (N=58)

The relationship is further explored through regression between the mean cumulative allocations and both the sum of draws and the order weighted sum of draws. Table 58 shows the slope coefficient and its associated p-value of a linear regression of the mean cumulative allocation and both the sum of draws and the order weighted sum of draws.

This shows that the regression slope for the sum of draws is statistically significant at the 10% level for each stage and is significant at the 5% level for many stages including the final stage. By contrast, the regression vs. the order weighted sum of draws for the later stages (stages 10-13) are not statistically significant at the 10% level. For no-feedback markets, the average cumulative allocations fit the sum of draws better than the order weighted sum of draws.

	Regression		Regression	
	Sum of Draws		Weighted Sum	
Stage	Slope	p Value	slope	p value
1	14.3	0.000	1.1	0.000
2	10.4	0.004	0.8	0.004
3	9.8	0.032	0.8	0.030
4	11.3	0.016	0.9	0.019
5	9.5	0.025	0.8	0.024
6	8.8	0.054	0.8	0.044
7	8.0	0.074	0.7	0.068
8	11.2	0.044	0.9	0.085
9	10.8	0.038	1.0	0.077
10	11.8	0.095	0.8	0.174
11	10.6	0.068	1.0	0.191
12	12.1	0.067	0.9	0.300
13	11.9	0.017	1.3	0.231

Table 58. Regression results for no-feedback markets by stage (e.g., number of participants).

Table 59 gives the same information for the feedback markets. Here the relationships have changed. In the later stages (stage 9-13) the relationship between the average cumulative allocation and the sum of draws are not statistically significant at the 10% level. However, for all stages, the average cumulative allocation and the order weighted sum of draws is statistically significant at the 5% level.

	Regression		Regression	
	Sum of Draws		Weighted Sum	
Stage	Slope	p Value	slope	p value
1	10.6	0.002	0.8	0.002
2	11.2	0.001	0.9	0.001
3	11.4	0.006	0.9	0.003
4	12.7	0.001	1.1	0.001
5	12.8	0.002	1.1	0.001
6	13.4	0.006	1.2	0.001
7	12.8	0.021	1.3	0.002
8	13.5	0.049	1.3	0.004
9	10.1	0.132	1.4	0.005
10	11.6	0.240	1.5	0.008
11	7.0	0.317	1.5	0.005
12	8.4	0.259	1.6	0.008
13	6.8	0.184	1.8	0.002

Table 59. Regression results for feedback markets by stage (e.g., number of participants).

4.4 Individual Allocations

The allocations by the participants are studied in this section.

4.4.1 Rounding and Position Bias

All allocations were examined for potential biases. Figure 19 is a histogram of the final digit of the allocation. For example, if the participant allocated 22 tickets on state B, then this chart would record that as a 2. In Figure 19 all allocations of 0 total have been removed (in other words, those shown as 0 are 10, 20, 30 etc.). This shows participants had a pronounced bias to round their allocations to a value a multiple of 10 or 5.

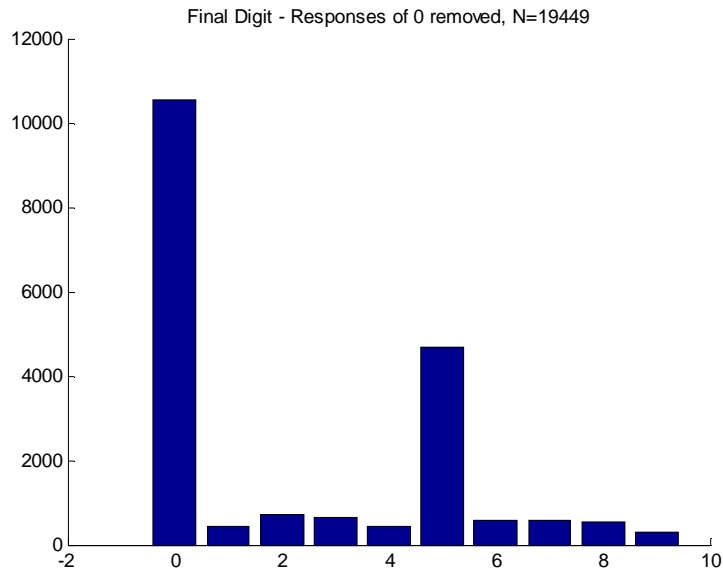


Figure 19. Histogram of final digit of allocation (e.g. 13 would be recorded here as 3) for all IP markets (with responses of 0 removed) illustrating the rounding effect

Since the predominant number of workers in AMT's workforce read left to right (as well as this is the way the AMT web pages and this survey were built), a tendency to allocate from left to right was investigated. Figure 20 shows two grouped histograms. First the ratio of tickets allocated to a particular state divided by the total allocation is shown (that is, the total of the distribution is 1). Along with this is the distribution of urn correct states. In the absence of any bias, these should be approximately the same. However they differ with a p-value of less than .001. Descriptively, notice that allocations to state A, B, C and D are all higher than expected and allocations to states E and F are both lower than expected. This supports the hypothesis of a left to right bias.

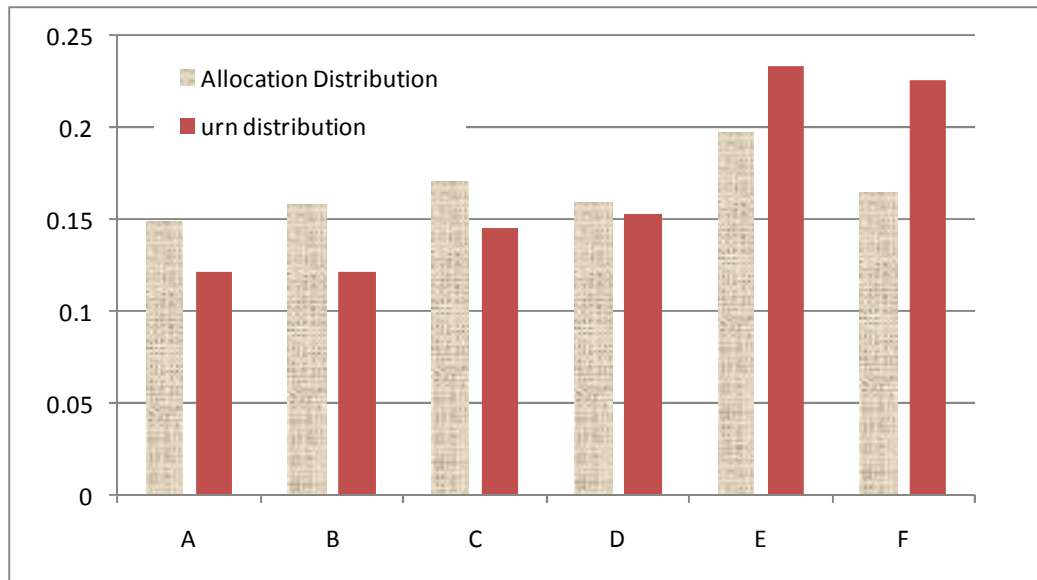


Figure 20. Comparison of average ticket allocation by state to urn correct draw

4.4.2 No-feedback Markets: Allocations vs. Probability

Figure 21 is a box plot of each individual allocation versus the corresponding Bayesian probability as derived from the three ball draws for the IP markets. The Bayesian probabilities are at discrete values because there are only three types of ball draws (all balls the same, two balls the same or all balls different) which breaks down to 7 different probabilities (see Table 22). The line where the allocation is equal to the Bayesian Probability has been drawn on the figure. Any allocation above this line could be considered an over allocation to the probability and any allocation below the line could be considered an under allocation to the probability. A least squares linear regression was performed and the resulting regression line is also drawn. Note that the regression line indicates that participants have a bias which creates over allocation for low probability situations and an under allocation for higher probability draws. One possible explanation for this effect is the well studied tendency for individuals to overestimate the probability of low probability events and underestimate the probability of high

probability events (section 2.2.2.2.2). As will be discussed later (section 5.1.3), this bias has a detrimental impact on market performance.

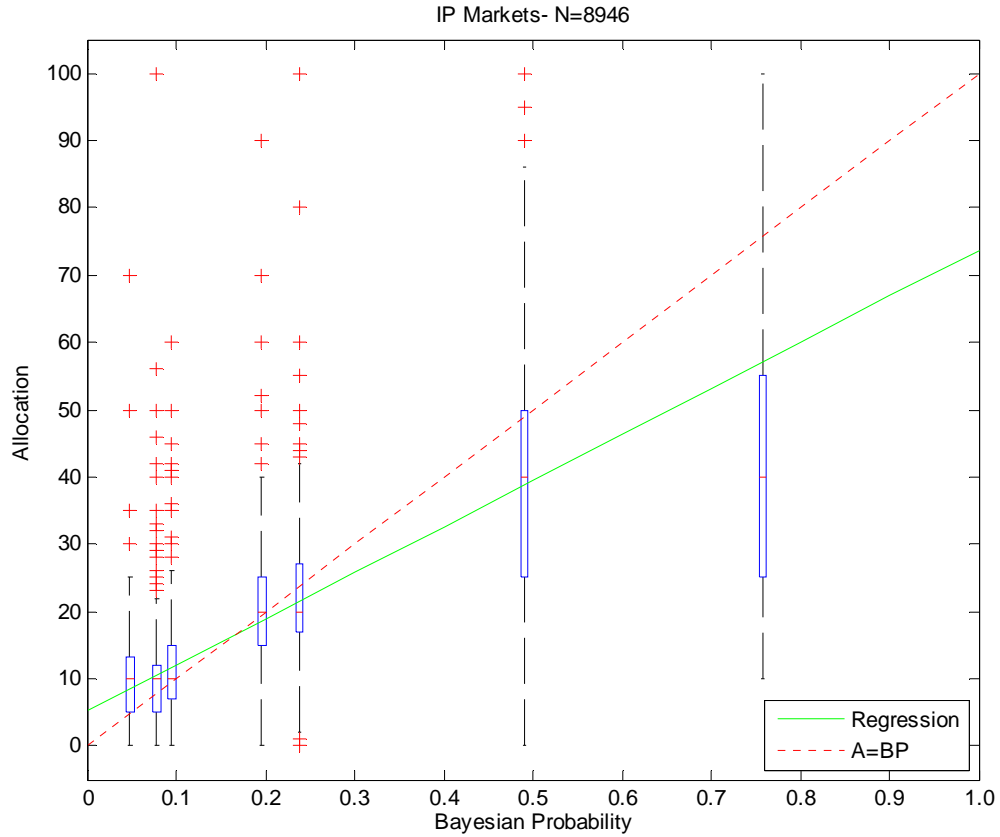


Figure 21. Allocations vs. Bayesian probability - IP markets. Data points are represented by the box plot. The dashed line shows where the Allocation is equal to the Bayesian Probability. The solid line is the regression fit to the data.

4.4.3 Feedback Markets: Influence of Public Data on Allocations

The influence of public information on the allocation by participants is studied. In this analysis the allocations are grouped by the number of balls drawn for a particular state.

To illustrate the method, consider an individual who has a ball draw of ECE and is presented a public cumulative allocation of: A:55; B:32; C:110; D:21; E:11; and, F:71.

This participant has been shown a total of 300 tickets. In terms of percent of this total

the cumulative allocations would be: A:18.3%; B:10.7%; C:36.7%; D:7.0%; E:3.7%; and, F:23.7%. For this illustration, imagine the participant allocated their own tickets as: A:4; B:3; C:40; D:3; E:40; and, F:10. Then for the groups the following data points would be added:

- 0 balls drawn group:
 - State A - Allocation:4; Public:18.3%
 - State B- Allocation:3; Public:10.7%
 - State D- Allocation:3; Public:7.0%
 - State F- Allocation:10; Public:23.7%
- 1 ball drawn group:
 - State C- allocation:40, Public:36.7%
- 2 balls drawn group:
 - State E- allocation:40, Public:3.7%
- 3 balls drawn group:
 - No data points added since this situation did not occur.

Each allocation can thus be categorized and scatter plots generated for each group.

Figure 22 shows a scatter plot for all allocations where one ball was drawn. The plots show the data for both frequency and percentage feedback markets. Note that this scatter includes responses for all stages of the market; so, in frequency markets participants were seeing different levels of total tickets allocated. For each scatter a least squares linear regression is calculated (using data from both frequency and percentage markets in the regression) and is shown as the line in the figure. As expected, the positive slope indicates that increasing levels of public allocation increase the participant's allocation.

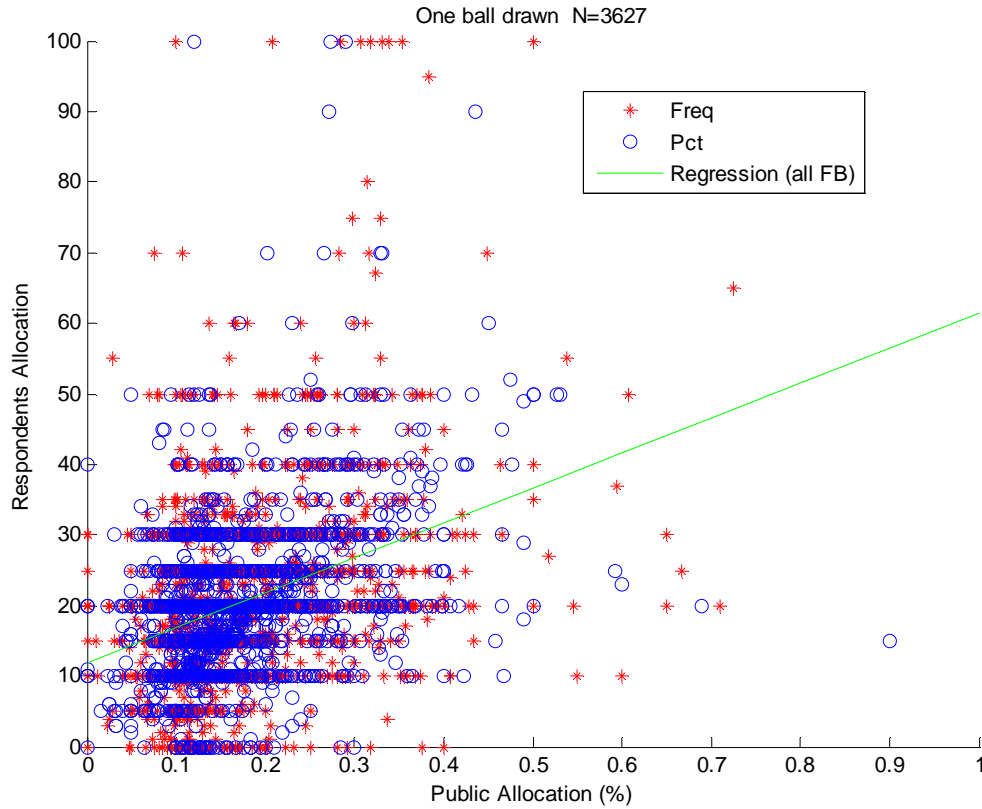


Figure 22. Scatter plot of participant's allocation vs. the public cumulative allocation for the case where they receive only one of the indicated ball in their draw set. The regression fit is shown as the line and is calculated using all data points (from both percentage and frequency markets).

The regression calculations for each of the four groups (number of balls drawn) are summarized in Table 60. The value of the slope is indicative of the strength of the effect of public data. Note that for 0-2 balls drawn the p-value of slope coefficient is 0 indicating the value is statistically significant at the 5% level. However, the R^2 value is relatively low for all cases. This indicates the relationship accounts for a very small percentage of the observed variance - at most 12.9% for the one ball drawn case. The Cook's distance (Cook, 1977) for the point with the maximum value is also recorded in the table. These are all comfortable levels as the general rule of thumb is a value above

1 is needed before a single data point can impact the overall regression values. Notable is the result for the 3 ball drawn case. Note that this case has the fewest observations with a sample size of 98. Here, the slope is not statistically significant at the 5% or 10% levels. Table 61 is included for reference, and shows the distribution of allocations under the same conditions for the no-feedback market case.

	Constant			Slope		
	B	SE	p	B	SE	p
0 balls drawn	5.89	.25	<.005	38.27	1.44	<.005
1 ball drawn	12.01	.41	<.005	49.51	2.14	<.005
2 balls drawn	23.70	1.44	<.005	50.24	6.38	<.005
3 balls drawn	47.31	6.73	<.005	-3.94	24.78	.874

	F	Max Cook's Dist	R ²	N
0 balls drawn	710.1	0.138	.096	6727
1 ball drawn	537.3	0.149	.129	3627
2 balls drawn	61.9	0.210	.067	870
3 balls drawn	.025	0.110	-.01	98

Table 60. Summary of regression analysis of the impact of public data on participant's allocations

Number of same state balls in draw set	Allocation (out of 100)		N
	Mean	SD	
0	10.42	6.83	5,901
1	21.48	9.39	2,578
2	39.13	19.52	1,006
3	44.43	26.16	61

Table 61. Summary of participant's allocations in no-feedback markets as a function of the number of a particular ball they received in their draw set.

Multivariate regression was preformed to understand the relative values of the influence of public data as the private information changes. A change in the slope would indicate a type of integration of public and private data in the sample. In the first regression analysis the public data is qualified by a dummy variable set to the number of ball draws.

The equation fit was:

$$\text{Predicted Allocation} = B0 + B1*\text{Dummy0}*(\text{Public in \%}) + B2*\text{Dummy1}*(\text{Public in \%}) + B3*\text{Dummy2}*(\text{Public in \%}) + B4*\text{Dummy3}*(\text{Public in \%})$$

where

$$\text{DummyN} = 1 \text{ if N balls drawn and } 0 \text{ otherwise}$$

The results of fit to this equation are shown below in Table 62. All coefficients are statistically significant and different. This result supports the conclusion that public and private information is being integrated. The values of the coefficients are seen to increase with increasing numbers of balls drawn. This indicates that as more informative private information is received, the participant is placing a higher value on public information.

	Coefficient	SE	p-value	VIF
B0	9.86	.24	<.0005	-
B1	17.51	1.50	<.0005	1.87
B2	59.49	1.49	<.0005	1.78
B3	104.95	1.94	<.0005	1.25
B4	120.85	4.23	<.0005	1.03

Table 62. Multivariate Regression Results. Fit to equation shown above. VIF (variance inflation factor) is a multicollinearity metric. Generally VIF values below 5-10 are considered to show low levels of collinearity.

In a second multivariate regression, the data is fit to a linear model with an interaction term. The equation fit is shown below:

$$\text{Predicted Allocation} = C0 + C1*\#_Balls_Drawn + C2*Public\ in\ \% + C3*\#_Balls_Drawn*Public\ in\ \%$$

Table 63 shows the values for the coefficients fit to the model above. The VIF (variance inflation factor - a multicollinearity metric) for the C3 term is slightly high at 6.25 indicating some of the variance it accounts for is shared by other coefficients. However, the value is less than 10 and does not substantially impact the model validity.

	Coefficient	SE	p-value	VIF
C0	5.34	0.30	<.0005	-
C1	8.344	0.34	<.0005	4.92
C2	39.99	1.69	<.0005	1.80
C3	5.28	1.625	.001	6.25

Table 63. Multivariate Regression Results. Fit to equation shown above. VIF (variance inflation factor) is a multicollinearity metric. Generally VIF values below 5-10 are considered to show low levels of collinearity.

The implied univariate regression coefficients derived from this multivariate analysis are shown in Table 64. These are similar to the values derived from the univariate regression as shown in Table 60. The notable exception is the implied relationship for 3 balls drawn, where the implied relationship has a smaller constant (30.37 vs. 47.31) and a large positive slope (55.84).

	Constant	Slope
0 balls drawn	5.34	40.00
1 ball drawn	13.68	45.28
2 balls drawn	22.03	50.56
3 balls drawn	30.37	55.84

Table 64. Linear relationships implied by the multivariate analysis.

After completing the markets, participants were asked "Did you encounter a situation where there was a contradiction between your private (draw) information and the market information?" A yes and no selection was presented. If the participant selected yes, then they were shown the choices displayed in Figure 23. 31% of the respondents indicated they attempted to integrate public and private data when they had conflicting inputs.

How did you resolve the contradiction?					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No conflict indicated	191	38.0	38.0	38.0
	I relied on the private information.	96	19.1	19.1	57.1
	I relied on the public information.	21	4.2	4.2	61.2
	I attempted to integrate the two information sources	156	31.0	31.0	92.2
	I just spread out my tickets across all options	39	7.8	7.8	100.0
	Total	503	100.0	100.0	

Figure 23. How participant's reported resolving conflicts between public and private data.

4.4.4 Categorizing responses into styles

Table 65 (random draws experiment) and Table 66 (mirage prone experiment) gives the results of fitting the experimental allocations to the models of different allocation styles (shown in Figure 12). The model compares the participant's allocation to each style and selects the one with the minimum Wurtz distance. If the minimum Wurtz distance is greater than 0.25 the response is grouped into the 'No Fit' category.

Analysis of the results in Table 65 show that for IP markets the dominant styles used are MaxiMin and DrawPct. DrawEV is seen to be relatively rare. The relative

relationship between DrawPct and DrawEV agrees with prior research where P, or probability, bets are observed more frequently than \$, or maximum expected value, bets (see section 2.2.3.2.3).

From Table 65, for feedback markets, the relatively large fraction which most closely matches the mimic style is surprising at 26.2% in percent markets and 22.1% in frequency markets. MaxiMin and Mimic, which are relatively low information and computation strategies, make up nearly 50% of the participants. Also surprising is the very low percentage of participant's allocations which match a strategy that attempts to integrate public and private data (AllPct, AllPctEV, AllDraws or AllDrawsEV). In fact there is a complete lack of participant allocations matching either the AllPctEV or AllDrawsEV strategies. The previous section (Section 4.4.3) showed that participant allocations were influenced by public information; however the effect must more subtle than the style models indicate.

	No	Percent	Frequency
	Feedback	Feedback	Feedback
MaxiMin	33.9%	24.4%	23.0%
DrawPct	48.1%	24.7%	24.2%
DrawEV	9.5%	7.0%	10.7%
Mimic	-	26.2%	22.1%
PublicEV	-	0.7%	1.0%
AllPct	-	9.4%	6.6%
AllPctEV	-	0.0%	0.0%
AllDraws	-	-	4.2%
AllDrawsEV	-	-	0.0%
No Fit	8.5%	7.6%	8.1%
Sample N	1,491	939	1,048

Table 65. Descriptive results from categorization of participant's allocations (from the random draws experiment)

	IP	Pct	Freq
MaxiMin	33.7%	23.6%	27.2%
DrawPct	40.3%	24.8%	24.8%
DrawEV	8.8%	6.7%	6.1%
Mimic		21.0%	17.5%
PublicEV		0.6%	0.5%
AllPct		11.1%	4.9%
AllPctEV		0.0%	0.0%
AllDraws			5.8%
AllDrawsEV			0.0%
No Fit	17.2%	12.2%	13.3%
Sample N	377	343	412

Table 66. Descriptive results from categorization of participant's allocations (from the mirage prone draw experiment - those failing attention test removed)

Table 67 gives a summary across three broad strategy types comparing MaxiMin, Percent based strategies (DrawPct+Mimic+AllPct+AllDraws) and MaxEV strategies (DrawEV+MimicEV+AllPctEV+AllDrawsEV) for the random draws experiment. The ratio of participants maximizing expected value is relatively stable in the range of 10% in each market type. Note that the MaxiMin fraction drops by nearly 10% when feedback

is present. Also, note that the fraction matching percent and EV types increases by approximately this same value.

	No	Percent	Frequency
	Feedback	Feedback	Feedback
MaxiMin	33.9%	24.4%	23.0%
Percent Type	48.1%	60.3%	57.2%
EV Type	9.5%	7.8%	11.7%
No Fit	8.5%	7.6%	8.1%

Table 67. Grouping of styles into major categories (random draws)

Table 68 shows the same summary for the mirage prone data set. While the percentage of responses which match to MaxiMin are similar in the two experiments, the percentage of No Fit is higher in the mirage prone draw experiment while the ratio of percent and EV type strategies is lower. The change in the bonus and base ratio (see Table 24) and the presence of a no-feedback market screener are likely reasons for the shift.

	No	Percent	Frequency
	Feedback	Feedback	Feedback
MaxiMin	33.7%	23.6%	27.2%
Percent Type	40.3%	46.4%	52.9%
EV Type	8.8%	7.3%	6.6%
No Fit	17.2%	12.2%	13.3%

Table 68. Grouping of styles into major categories (mirage prone draw)

Participants' responses can also be grouped in terms of information usage. Table 69 is a recasting of Table 65 in terms of information usage. MaxiMin falls into the 'Little Info Use' category. DrawPct and DrawEV fall into the 'Primarily Private Info' category. Mimic and PublicEV fall into the 'Primarily Public Info' category. AllPct, AllPctEV, AllDraws and AllDrawsEV fall into the 'Integrate Public & Private' category. In no-

feedback markets approximately one third of participants' responses fit to models which imply little use of information, while approximately 58% fit to models which did use the private information available. In feedback markets approximately a quarter of participants' responses were fit to models which use little information (e.g. MaxiMin). Approximately one third fit to models which use only private information (e.g. allocation in proportion to probabilities). Another quarter fit to models which use only public information (e.g. allocating in proportion to public information). Finally, only 10% of participant's responses fit to models which integrate public and private information.

	No	Percent	Frequency
	Feedback	Feedback	Feedback
Little Info Use	33.9%	24.4%	23.0%
Primarily Private Info	57.6%	31.7%	34.9%
Primarily Public Info	-	26.9%	23.2%
Integrate Public & Private	-	9.4%	10.8%
No Fit	8.5%	7.6%	8.1%

Table 69. Categorization of participants' responses by use of information (random draws experiment)

5 Discussion, Conclusions and Next Steps

This section discusses and interprets the results, suggests future work and analyzes the application of simulation technology to the study of markets.

5.1 Conclusions

The research conclusions are reviewed in the context of the research questions and experimental hypotheses.

5.1.1 What are the Mechanisms Involved in the Information Assimilation Process?

The idea and understanding that markets can assimilate information dates back to some of the earliest work on economics. Smith (1776) and Hayek (1945) wrote of this and the concept was used to build the fundamental *Theory of Rational Expectations* and the *Efficient Market Hypothesis* - important pillars in modern economics. The ability of markets to assimilate information was empirically shown by Plott (1988). In the present study the process of assimilation was discussed in terms of two components:

- The collection of diverse information - individuals using their own information make a judgment about the implications of that information and bring this to the market.
- The market's judgment on the implications of this gathered information - an inductive process which Plott (2003) and others call information aggregation.

By studying markets with and without feedback these two mechanisms can be studied in more detail.

Hypothesis:

1. Markets without feedback can collect information and outperform the average of their parts.

Hypothesis supported - In section 4.2.1 it was shown that no-feedback markets are capable of collecting information in that their performance exceeded that of the collection of the individuals. Specifically, the null hypotheses that Decisiveness and Accuracy were equal for markets and the collections of individuals were both rejected (Table 30).

Hypothesis:

2. Feedback markets have an inductive process, above and beyond collection, which creates agreement (precision).

Hypothesis supported - Mean entropy of the cumulative allocations was examined. The data showed that feedback markets had statistically significantly lower entropy than markets without feedback; this is an indication of the inductive information aggregation mechanism.

Hypothesis:

3. Feedback markets have an inductive process, above and beyond collection, which can assimilate more information than markets without feedback.

Hypothesis supported - In section 4.2.2 the performance of markets with and without feedback was explored in order to understand the additive impact of the market's inductive processes. The performance of the markets in terms of their decisiveness, the ability of a market to come to a statistically significant conclusion, and correctness, the ability of a market to come to a correct conclusion, was examined. It was shown that getting aggregation levels beyond the no-feedback case is a tradeoff with feedback markets offering additional decisiveness at the expense of correctness. The actual induction mechanism itself was examined by looking at the mean Wurtz distance between the cumulative allocations and the AIA distribution (or Aggregate Information Available, the Bayesian posterior probability of the simple sum of all ball draws - a potential measure for perfect complete information aggregation as suggested by Plott et al (2003)). The data showed that feedback markets' cumulative allocations had statistically significantly lower Wurtz distances to the AIA distribution indicating they were aggregating information above and beyond the collection of information by no-feedback markets.

Hypothesis:

4. Information mirages are present in markets with feedback. They are an artifact of the inductive process. These mirages are an error mechanism which reduces the market's overall accuracy.

Hypothesis supported - Information mirages were postulated to be one of the failure mechanisms of feedback markets. In the random draws experiment some clues were

observed which supported information mirages as a mechanism. Section 4.2.3.1 reviewed the allocations of several feedback markets in detail which provided some qualitative evidence for information mirages. In section 4.2.3.2 the metric 'Sum of AIA distances' (section 3.4.1) for each market in the random draws experiment was compared to the market's aggregation performance by looking at the Wurtz distance between the current AIA distribution and the cumulative ticket allocation. A statistically significant linear relationship (with positive slope) was seen for feedback markets, while no such relationship was seen for no-feedback markets.

The information mirage phenomenon was studied further using a mirage prone draw set which was repeatedly applied to all markets. This was seen to negatively impact the correctness of both no-feedback and feedback markets; but with a much greater impact to the feedback markets as shown in Table 70. After 13 stages, the correctness of feedback markets was seen to drop to 28.6% compared to 74.6% in the random draws experiment. This change was statistically significant at the 5% level (p-value of less than .001). The correctness of the no-feedback markets was only reduced from 81.5% to 65%. This change was not statistically significant at the 5% level (p-value of .173). Note that there were other factors which may have contributed such as the change in incentives and the use of screening methods in the random draws experiment (see Table 24).

		No-feedback	Feedback	p-value
Decisiveness	Random Draws	56.3%	70.8%	0.065
	Mirage Prone	60.6%	48.3%	0.180
	p-value	.436	.005	
Correctness	Random Draws	81.5%	74.6%	0.338
	Mirage Prone	65.0%	28.6%	0.013
	p-value	.173	<.001	

Table 70. Comparison of decisiveness and correctness of no-feedback and feedback markets at stage 13 over the two experiments.

The fingerprint of information mirages was seen in several analyses of the mirage prone data set. In section 4.3.2.1 it was seen that in contrast to the result of the random draws experiment, the Wurtz distance between the AIA distribution and the final cumulative allocation was statistically significantly lower for no-feedback markets. A histogram (Figure 17) showed that the feedback markets had long tails of very low values - where induction was creating agreement on correct assessments - and very large values - where induction was creating agreement on incorrect assessments as information mirages. Similarly, the distributions for entropy were studied and also shown to have statistically significant greater variance.

The most conclusive evidence for information mirages in the feedback markets was seen in section 4.3.2.3. The relationship between the mean number of tickets allocated and the sum of draws and the order weighted sum of draws was investigated. Table 58 showed a statistically significant linear relationship between the sum of draws and the cumulative allocation for no-feedback markets. By contrast, the relationship between the order weighted sum of draws and cumulative allocations was not significant at the end of the market. In stark contrast, Table 59 showed the opposite. For feedback markets the

relationship between the sum of draws and cumulative allocations was not significant at the end of the market while the relationship between the order weighted sum of draws and cumulative allocations was statistically significant. This indicates that for feedback markets early draws have a persistent impact on the cumulative allocation to later participants.

Charles Mackay, in his book ‘Extraordinary Popular Delusions and the Madness of Crowds (1932)’ states:

"Men, it has been well said, think in herds; it will be seen that they go mad in herds, while they only recover their senses slowly, and one by one."

He conveys a perspective that the root causes of bubbles are drastically *irrational* behavior brought about by the situation of the interactions. However, another perspective is that *irrational* behavior is not required to generate bubbles. Is the logical argument of the third person in an information cascade whose preceding participants have both chosen a different door than their private information indicates (as discussed in section 2.3.2.1) irrational? Instead, bubbles and mirages can also be caused by system structure, diversity in participants and probabilistic variations in information. Vives writes (2008):

"It is found that apparently contending theories, such as market informational efficiency and herding, build in fact on the same principles of Bayesian decision making. The upshot is that we do not need "irrational" agents to explain herding behavior, crisis, and crashes."

In fact, a market subject to a mirage prone data resembles a Prisoner's Dilemma (Zwick, 2011) where individual rationality results in collective irrationality.

Table 71 summarizes the experimental hypotheses and their evaluation in this work for the research question "What are the mechanisms involved in the information assimilation process?"

Hypothesis	Evaluation
1. Markets without feedback can collect information and outperform the average of their parts.	Supported - Section 4.2.1 - Markets vs. the collection of individual allocations have statistically significant higher decisiveness and correctness.
2. Feedback markets have an inductive process, above and beyond collection, which creates agreement (precision).	Supported - Section 4.2.2.3 - Feedback markets have statistically significantly lower entropy.
3. Feedback markets have an inductive process, above and beyond collection, which can assimilate more information than markets without feedback.	Supported - Section 4.2.2.2 - Feedback markets have statistically significantly lower Wurtz distances between the cumulative allocations and the AIA distribution.
4. Information mirages are present in markets with feedback. They are an artifact of the inductive process. These mirages are an error mechanism which reduces the market's overall accuracy.	Supported - Section 4.2.3.1 showed qualitative data on the presence of information mirages. Section 4.2.3.2 showed that feedback markets have a statistically significant linear relationship between the Wurtz distance between the AIA distribution and the cumulative allocation and the sum of Wurtz distance between the intermediate (stage) AIA and the final AIA distribution. Section 4.3.1.1 showed a dramatic and statistically significant reduction in both decisiveness and correctness for feedback markets when the mirage prone draw set was applied. By contrast, no-feedback markets showed no statistically significant change. Section 4.3.2.1 showed that under the mirage prone draw set the no-feedback markets exhibited statistically significantly smaller Wurtz distance between the cumulative allocation and the AIA distribution than feedback markets. This was seen to be the exact opposite relationship observed in the random draws experiment. Section 4.3.2.3 showed a statistically significant linear relationship between the sum of draws and the cumulative allocation for no-feedback markets (and no relationship for feedback markets). By contrast, a statistically significant linear relationship between the order weighted sum of draws and the cumulative allocation was found for feedback markets (and no relationship for no-feedback markets).

Table 71. Summary of hypotheses and evaluations for the first research question - "What are the mechanisms involved in the information assimilation process?"

5.1.2 How does the Presence and Type of Feedback Impact Market Performance?

Hypothesis:

5. Adding feedback increases the decisiveness of a market at the expense of correctness.

Hypothesis Supported - Section 4.2.2.1 showed that markets with feedback had statistically significantly higher decisiveness than markets without feedback. However, markets without feedback had statistically significantly higher levels of correctness.

Hypothesis:

6. The type of feedback can modulate the induction process - more informative feedback increases decisiveness at the cost of correctness.

Hypothesis Supported - Section 4.2.2.1 showed that frequency feedback markets had statistically significantly higher decisiveness than percent feedback markets.

Descriptively, percent feedback markets had higher correctness than frequency feedback markets, but that difference was not statistically significant at the 5% level.

The addition of feedback and the choice of type of feedback affords a tradeoff between decisiveness and correctness. Figure 24 (duplicate of Figure 13 shown here for the reader's convenience) illustrates this tradeoff. Adding feedback increases the information assimilation ability of a market by adding the inductive ability. This increases decisiveness and reduces entropy. However, it brings the possibility of

information mirages which reduce correctness. Frequency feedback, which delivers more information to the market than percent feedback, descriptively increases both effects over percent feedback. This suggests there is the possibility to modulate the inductive effect to some extent and match the market performance to the needs of the organizer (see section 5.3.1 for more discussion).

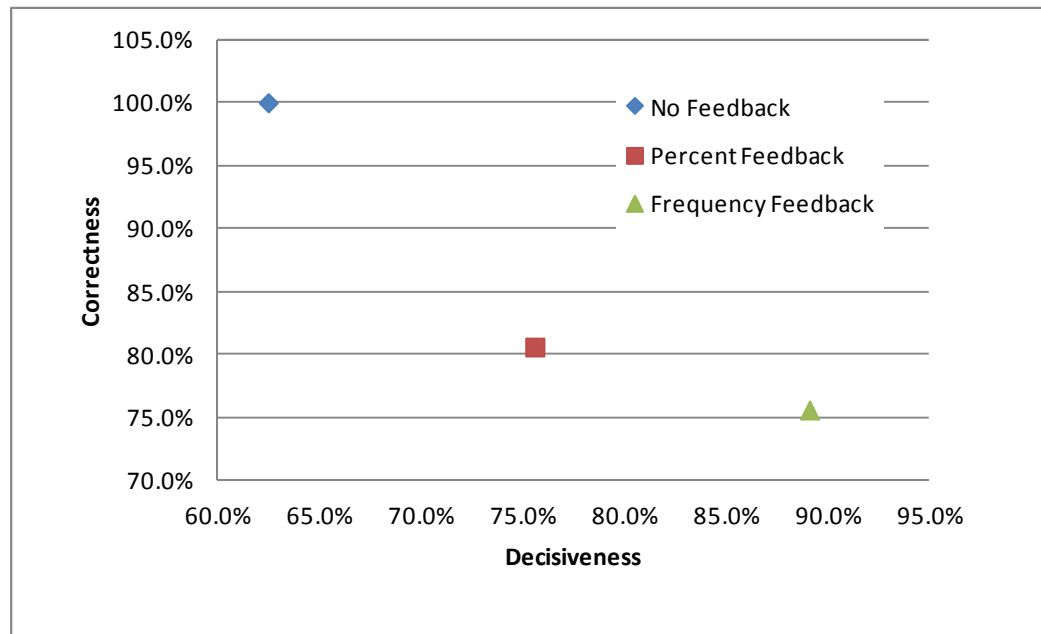


Figure 24. Summary of market performance illustrating the tradeoff between correctness and decisiveness after stage 20 of the random draws experiment.

One analogy is to think of the feedback market as an amplifier - a Bayesian amplifier. An amplifier takes a weak signal and turns it into a strong signal. Unfortunately, amplifiers will amplify noise just as well as signal. Changing the presence and type of feedback could be thought of as changing the amplification level.

In section 1.1.3 one of the four conditions in Surowiecki's main thesis for the wisdom of crowds was independence. Complete independence would imply no feedback. In fact,

no feedback markets did have distinct advantages, such as higher correctness. However, markets with feedback, and thus some level of dependence, also displayed valuable characteristics, such as higher decisiveness.

Comparing the results of section 4.2.2.2, where it was shown that markets with feedback had lower Wurtz distances to the final AIA distribution, and section 4.2.2.1, where it was shown that no feedback markets had higher correctness, presents an apparent paradox.

Both of these results represent measures of accuracy and feedback markets outperformed in the first and no-feedback markets outperformed in the latter. The resolution to the paradox comes from realizing that correctness is based on the two stage evaluation process described in section 3.1.1.5. Once a market allocation is judged as correct, additional allocation of tickets to the correct state will not increase the measures of correctness; however, they will decrease the Wurtz distance to final AIA. So, once statistical significance of the result is established, additional gains from the inductive process do not benefit correctness. However, this inductive process does create information mirages which are adversely affecting correctness. This explains how information aggregation is simultaneously increasing accuracy as judged by the Wurtz distance to final AIA and decreasing accuracy as judged by correctness.

Table 72 presents a summary of the research hypotheses and their evaluation for the research question "How does the presence and type of feedback impact market performance?"

Hypothesis	Evaluation
5. Adding feedback increases the decisiveness of a market at the expense of correctness.	Supported - Section 4.2.2.1 showed that markets with feedback had statistically significantly higher decisiveness than markets without feedback. However, markets without feedback had statistically significantly higher levels of correctness.
6. The type of feedback can modulate the induction process - more informative feedback increases decisiveness at the cost of correctness.	Supported - Section 4.2.2.1 showed that frequency feedback markets feedback had statistically significantly higher decisiveness than percent feedback markets. Descriptively, percent feedback markets had higher correctness than frequency feedback markets but that differences was not statistically significant at the 5% level.

Table 72. Summary of second research hypotheses and their evaluation "How does the presence and type of feedback impact market performance?"

5.1.3 How do Individuals use Information in the Allocation Process?

Hypothesis:

7. In the absence of public data, participants allocate tickets in a Bayesian-like method (with base rate neglect growing with increasing number of matching balls drawn)

Hypothesis partially supported - The literature, as discussed in Section 2.2, presented potentially conflicting arguments for non-Bayesian behavior. For example, Base rate neglect (section 2.2.2.2.1) and pseudodiagnosticity (section 2.2.2.2.3) might suggest a different behavior than conservatism (section 2.2.2.2.2).

The participants tended to behave more as predicted in section 2.2.2.2.2.. Figure 21 in section 4.4.2 illustrated that participants over-allocate in low Bayesian probability circumstances and under-allocate in high Bayesian probability circumstances

In the case of markets, this misallocation tendency has a negative effect on market performance as well as the individual's financial performance. Figure 25 shows the same data with an overlay to explain this. As will be shown in the simulation in section 5.2.5 markets where participants use more computation and information intensive strategies (such as maximizing expected value) improve market performance. This implies that allocations which would move the regression line closer to the allocation is equal to the probability line would result in improved market performance. Further, a strategy that maximizes expected value will result in a regression line that under allocates on low probability events and over allocates on high probability events.

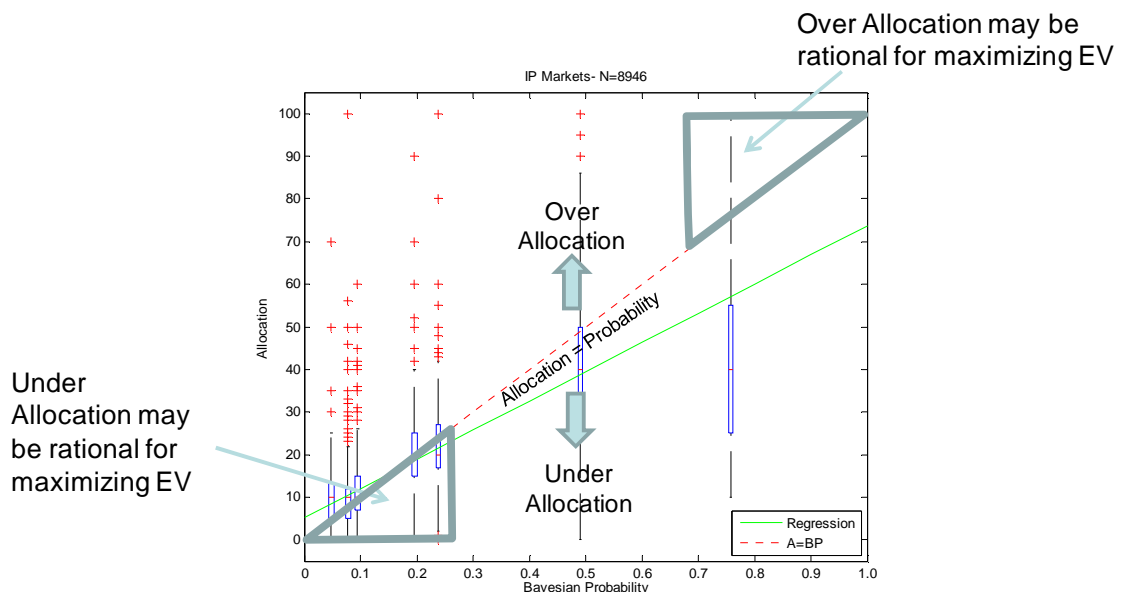


Figure 25. Allocations vs. Bayesian probability from no-feedback markets.

Hypothesis:

8. In the presence of public data, participants allocate tickets by integrating public and private data

Hypothesis Supported - Section 4.4.3 analyzed the relationship between participants' allocations, the public information available and their private information. Multivariate regression indicated as the number of balls drawn for a particular state increased in the draw set the influence of public data increased. Table 63 showed the interaction term (C3), which attempts to directly assess the integration effect, to be statistically significant and positive in value. The analysis with coefficients shown in Table 62 also demonstrated a differing influence of public data depending on the number of balls drawn. This analysis, however, did not constrict the model to linearity as the previous model did. The increase in the coefficients demonstrated a non-linear effect. Table 73 lists the coefficients from Table 62 along with the value of the increase. Notice that the rate of change in the coefficient values increases and then decreases as the number of balls drawn increases.

Coefficient	Value	Difference from prior Coefficient
B1 (0 Balls Drawn)	17.51	-
B2 (1 Ball Drawn)	59.49	41.98
B3 (2 Balls Drawn)	104.95	45.46
B4 (3 Balls Drawn)	120.85	16.00

Table 73. Coefficients from Table 62

The results for multivariate regression and univariate regression were different for the case where three balls were drawn. Table 60 showed the univariate analysis to predict no relationship between the allocation and public data. However, the multivariate analysis did predict a relationship regardless of the equation form (Table 62 and Table 63). Figure 26 shows the scatter plot of allocations versus public allocation for the three balls drawn case. The univariate (from Table 60) and multivariate (from Table 63) predicted allocation lines are shown. Visual inspection shows neither line to be an obviously superior fit to the data. The small sample size ($N=98$) for this case is likely to be part of the issue for the univariate model. In the multivariate case, the model is able to draw upon the other cases to strengthen the overall model and thus may have superior prediction capability.

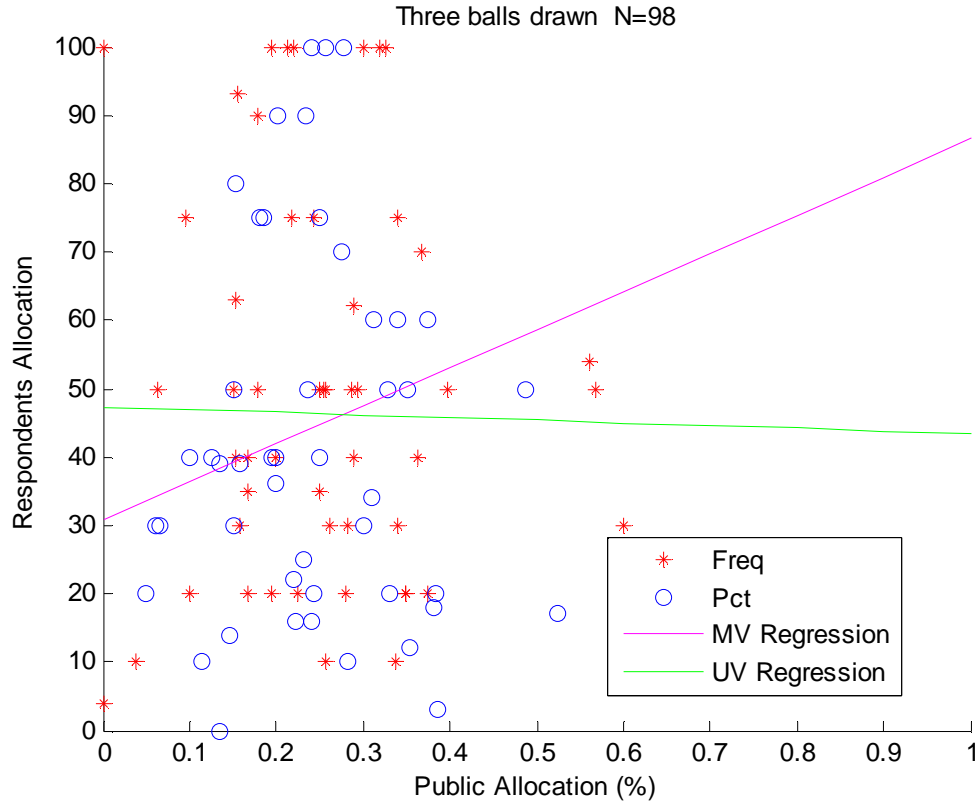


Figure 26. Allocations vs. public data for 3 balls drawn.

In addition to non-linearity in the relationship between allocations and balls drawn, there may be non-linearity in the relationship to allocation and public data. Figure 27 shows the allocations by participants in feedback markets for the random draws experiment when two balls were drawn. The pluses are allocations made by the AllPct algorithm (section 4.4.4). The pluses vary around a continuous curve because the allocation to an individual state has dependence on the distribution of the rest of the public allocations and how these relate to the ball draw in the draw set. The pluses trace a non-linear curve with changes in public allocation having more impact at lower values of public allocation. The slope of the regression fit to the experimental data points appears

similar to the slope of the reference curve for higher values of public allocation. There is some resemblance of the AllPct style curve and the shape of the experimental data points; however, the experimental points are notably lower (that is lower allocation than the AllPct style curve).

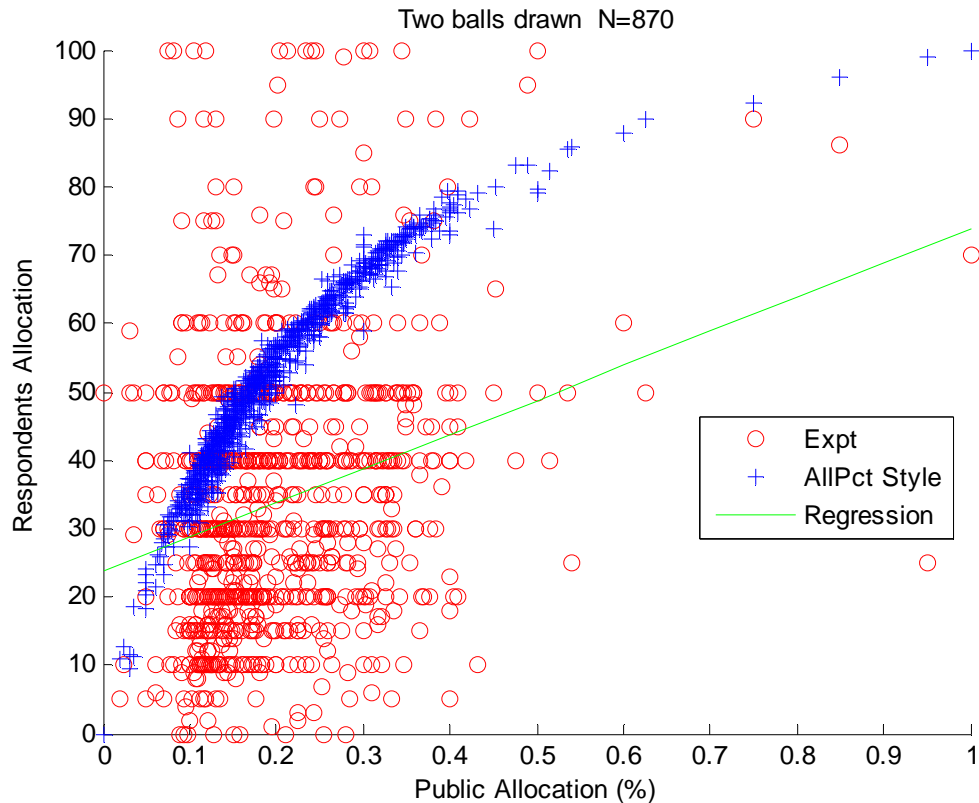


Figure 27. Participant's allocation vs. public allocation from random draws experiment for two balls drawn (shown as circles) with reference points added for the AllPct strategy (shown as pluses).

Hypothesis:

- Participants will favor bets with a higher probability of winning over bets which maximize expected winnings

Hypothesis Supported - Section 4.4.4 showed that strategies that maximize expected value are quite rare while strategies that generally allocate to percentages are common. One potential explanation for this behavior can be derived from analysis using the Prospect Theory curve of values.

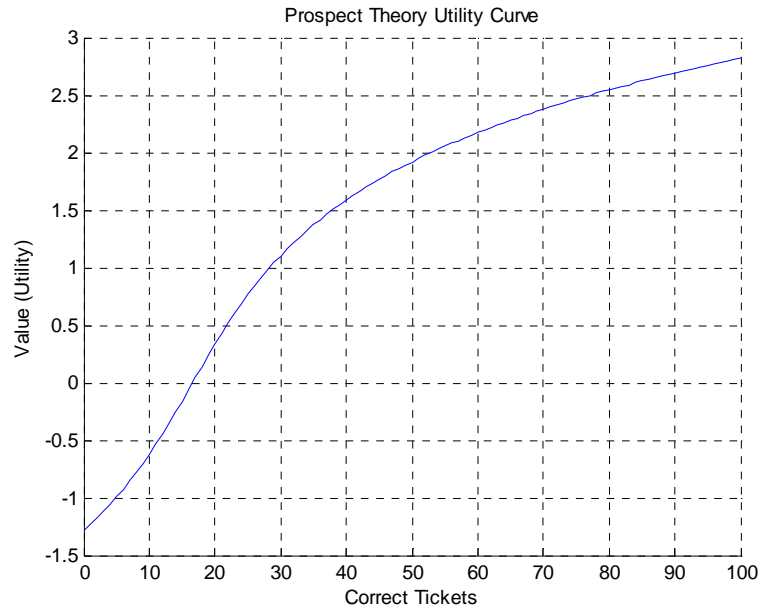


Figure 28. Simulated Prospect Theory Utility Curve

Figure 28 shows a representation of a participant's utility curve. The curve was generated with the equation

$$Value = \sinh^{-1}\left(\frac{Correct\ Tickets}{10} - \frac{16.6}{10}\right)$$

The inverse of sinh was used simply because it's shape resembles published Prospect Theory curves. The curve is shifted such that 0 utility is achieved for 16.6 correct tickets. This is chosen because a participant is guaranteed 16.6 correct tickets on average simply

by spreading their allocations uniformly. Expected utility can then be computed for various allocations and these can be searched for a maximum. To illustrate, consider a no-feedback market where a participant's draw was AAB. To maximize the dollar payout, a participant should allocate 100 tickets on A. The expected payout is

$$100*(.49)+0*(.51)=49 \text{ correct tickets}$$

Given the utility curve shown, the expected value of this allocation is:

$$\begin{aligned} &\text{Value}(100 \text{ correct})*(.49) + \text{Value}(0 \text{ Correct})*(.51) \\ &2.82*.49 + (-1.28)*.51 = 0.73 \end{aligned}$$

Using numerical analysis, a maximum for this was found for 70 tickets on A, 30 tickets on B and 0 tickets on the remaining. Its utility is

$$\begin{aligned} &\text{Value}(70 \text{ correct})*.49 + \text{Value}(30 \text{ correct})*.196 + \text{Value}(0 \text{ correct})*.314 \\ &2.36*.49 + 1.04*.196 + (-.853)*.314 = 1.09 \end{aligned}$$

So, given this utility curve, in order to maximize the expected utility for this draw, the participant would not allocated all 100 tickets on A. Of course this utility curve was not derived from any empirical data and was only used to illustrate the point, but the analysis does show a rational explanation for the small fraction of participants whose allocations matched the strategies that maximize the expected payout.

As noted, analysis of the individual responses indicate evidence for information integration. In section 4.4.4 categorization of individual responses in feedback markets showed approximately 10% of participants' responses fit to models which integrate public and private data. In section 4.4.3 multivariate analysis showed a statistically significant evidence for information integration taking place. However, these effects are rather subtle compared to the resounding evidence for information integration seen in the market results. This illustrates that the effects seen in markets represent an emergent phenomenon (Gilbert & Troitzsch, 2005). That is, the market results are a phenomenon that might not be predicted by analyzing the behavior of individuals in different market situations. For future work, it would be interesting to study the allocations by individuals in different feedback situations. A heterogeneous regression analysis in this manner could uncover new perspectives in the information integration process. It is quite possible that the 90% of respondents whose responses do not fit to a model which integrate public and private data may have a subtle bias in their responses which do integrate data but are not strong enough to register as that type of algorithm.

Table 74 presents a summary of the hypotheses and their evaluation for the research questions "How do individuals use information in the allocation process?"

Hypothesis	Evaluation
7. In the absence of public data, participants allocate tickets in a Bayesian-like method	Partially Supported - Section 4.4.2 showed that participants over allocate in low Bayesian probability circumstances and under allocate in high Bayesian probability circumstances.
8. In the presence of public data, participants allocate tickets by integrating public and private data	Supported - In section 4.4.3 multivariate regression indicated as the number of balls drawn for a particular state increased in the draw set the influence of public data increased.
9. Participants will favor bets with a higher probability of winning over bets which maximize expected winnings	Supported - Section 4.4.4 showed that strategies that maximize expected value are quite rare while strategies that generally allocate to percentages are common.

Table 74. Summary of hypotheses and their evaluation for the third research question "How do individuals use information in the allocation process?"

5.2 Simulating Markets

Simulation is capable of bringing valuable insights to the study of markets. The idea of simulating markets is explored in this section and a prototype simulator to test these concepts is explored.

5.2.1 Simulation Goals

The experimental methods used in the present research have been shown to be effective in studying market performance under different conditions. However, developing the software to carry out such research is complicated, collecting data from live participants can be time consuming and participants must be compensated. Building computer

models to simulate markets is an alternative method which allows investigation of various market characteristics without these limitations.

Simulation models can also be used to understand phenomenon seen in real world markets. Epstein and Axtell (1996) argue that agent based simulation provides a different way to explain social phenomenon . They argue that being able to explain a phenomenon is equivalent to asking if one can 'grow' it in an artificial environment. They express this process as “a generative kind of social science.”

A simulation model is a representation of reality and can never fully recreate the effects seen in the real world. In simulating a market, as in any simulation project, it is critical to identify the specific goals of the simulation in order to tailor and optimize the simulation towards those goals (Sterman, 2000).

5.2.2 Characteristics of Interest

A simulation platform can be used to study various characteristics of a market. Table 75 list just a few factors and methods to study them in the simulation environment. Simulation allows study of these factors by conducting extensive sensitivity analyses.

Factor	Simulation Procedure
Number of participants	Simulate over a range of participants and vary by market type to understand the sensitivity
Diversity	Change the ratios of allocation strategies to determine the impact on performance
Degree of positive feedback	Simulate a varying percentage of participants whose allocation strategies exhibit positive feedback
Degree of negative feedback	Simulate a varying percentage of participants whose allocation strategies exhibit negative feedback
Signal Strength	Simulate with varying signal strength. In this case signal strength is increased by adding more balls of the correct state to the urn.
Noise Level	Noise can be simulated by randomly altering a ball draw

Table 75. Summary of factors which can be studied in a market simulator.

The performance of the market as a function of the number of participants in the market is of great interest to the market designer. Depending on the circumstances it is likely under some degree of control by the market implementer. For example, the designer may wish to understand the market performance with few participants.

Other factors of interest may be more difficult for the designer to influence. The degree of diversity, the degree of positive feedback and the degree of negative feedback are key examples. The designer may influence diversity by who they invite to participate in the market. Training for participants could impact the positive and negative feedback effect. For example, helping participants to understanding the cascade tendency could decrease positive feedback.

Other factors may be outside the control of the designer but having an understanding of these factors could help in the analysis of the market results. Key factors in this

category are signal strength and the presence of noise. Signal strength relates to the amount of information available to market participants in the problem space. Noise relates to how clearly available information is interpreted. To illustrate, consider a forecasting prediction market. If the participants have a lot of information relevant to the forecast (e.g., inventory levels, customer insights, ecosystem factors, etc.) they would be able to make more knowledgeable assessments. If the information is unclear (e.g., customers giving conflicting signals, channel partners giving ambiguous responses, etc.) or noisy then individual assessments are more difficult.

5.2.3 Simulator Architecture

Figure 29 shows the architecture of the prototype prediction markets simulator. The blocks include:

Set of Allocation Strategies - A set of allocation strategies are needed. In the prototype simulator the strategies described in section 3.4.2 are used. Different types of markets are defined by the range of strategies used in the market. For example, for a no-feedback market (Incentivized Poll) only strategies that use private data are employed. For percentage feedback markets, only strategies that use percent values of public data are used.

Select Strategy - Once a set of possible strategies are defined based on the market type to be simulated one strategy from this set is chosen using a random process. The selected strategy is then passed into the Generate Allocation block.

Generate Allocation - This block receives a strategy from the select strategy block, a draw from the Draw Engine and public information from the Market block. When the allocation is generated it is delivered to the Market block.

Draw Engine - This block simulates the two urn processes. It first selects a ball at random as the correct state and then populates a second urn with a biased distribution of balls. This second urn is then sampled to generate the draw set.

Market Block - The Market block is a collection of all previous allocations and maintains the current cumulative allocation.

Stochastic Generator - This block is used to simulate randomness. Two sources of randomness were simulated: 1) noise in evaluating the draw; 2) noise in generating the allocation. In the draw evaluation noise process one or more of the draws delivered from the draw engine are randomly changed. A model parameter is used to vary the degree of this noise. The second source of randomness involves random deviation from the strategy derived allocation. In the implementation used some number of tickets are added to, or subtracted from, the allocation generated by the allocation strategy. The level of this degree of randomness is controlled by another parameter.

Simulation Controller - the controller block repeats the sequence until the specified number of market stages are complete.

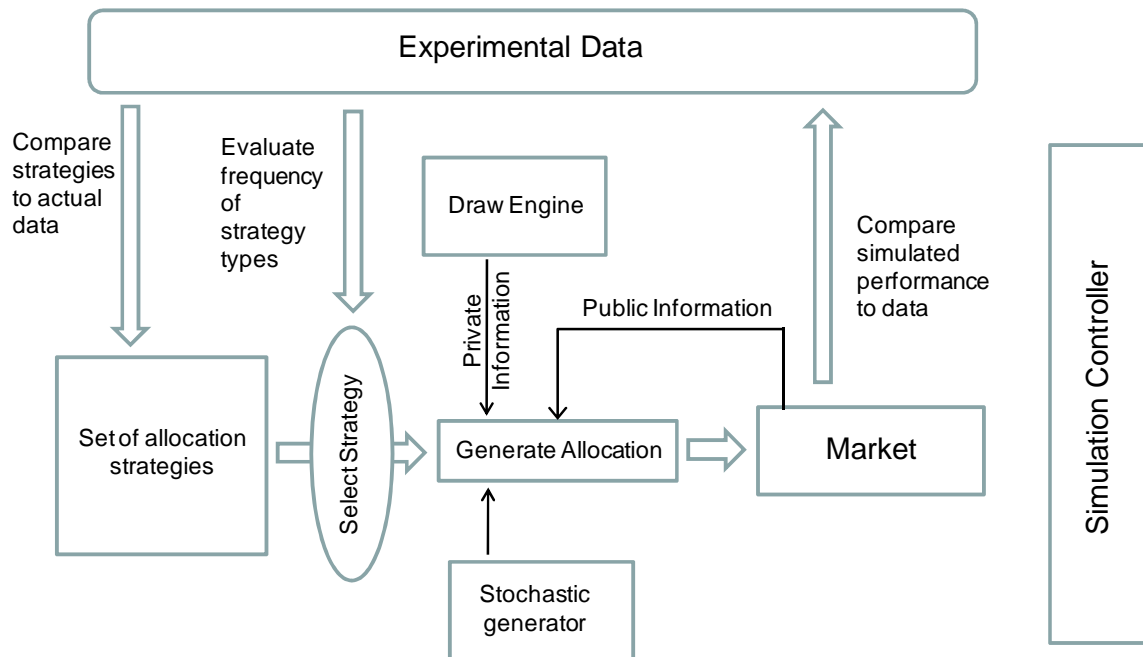


Figure 29. Market Simulator Architecture

5.2.4 Simulator Calibration and Validation

Once the architecture was implemented the parameters of the simulator must be specified; that is, the model had to be calibrated. The set of allocation strategies described in section 3.4.2 were used. The simulator used the Matlab command *randsample* to select one strategy from the alternatives. This method allows sampling with replacement based on a vector of probabilities. These probability distributions used were based on the frequency of each strategy observed from the experimental data as discussed in section 4.4.4.

Sets of market simulations were compared to experimental data in order to validate the models. Two methods were used in this process. First, the raw market results such as number of correct, incorrect and not significant were compared to experimental data.

Second, individual allocations generated by the simulator were used to extract allocation style characteristics as defined in section 4.4.4. These were then compared to the experimental data. The parameters in the stochastic generator were then adjusted to improve the fitness of the model generated data in relation to the experimental data.

The distributions across styles shown in Table 65 were used for the initial simulation. Values of the stochastic engine parameters were varied to improve the fit. Table 76 and Table 77 show a comparison of the experimental and simulated decisiveness and correctness for the prototype. A Fisher's Exact test p-value testing the differences is also shown. The goal of the calibration process was to make the differences not significant. The simulator does a good job of matching the decisiveness and correctness of the no-feedback markets with p-values usually greater than .05. The match for feedback markets is not as good, especially for the latter stages of the market.

Stage	No Feedback			Feedback		
	Experiment	Simulation	p value	Experiment	Simulation	p value
1	33.3%	37.8%	0.328	36.0%	22.7%	0.005
2	43.8%	49.0%	0.294	38.2%	37.2%	0.468
3	39.6%	49.0%	0.136	48.3%	45.3%	0.331
4	35.4%	49.0%	0.049	47.2%	49.5%	0.380
5	39.6%	49.8%	0.115	57.3%	54.3%	0.333
6	52.1%	50.4%	0.472	64.0%	55.8%	0.081
7	60.4%	53.0%	0.203	66.3%	57.6%	0.068
8	56.3%	54.6%	0.475	69.7%	58.7%	0.027
9	58.3%	55.4%	0.408	62.9%	61.1%	0.414
10	62.5%	55.4%	0.214	65.2%	62.8%	0.374
11	58.3%	59.0%	0.522	68.5%	61.7%	0.122
12	58.3%	62.2%	0.352	69.7%	62.0%	0.093
13	56.3%	62.8%	0.229	70.8%	62.5%	0.074
14	58.3%	63.0%	0.312	71.9%	63.7%	0.074
15	56.3%	64.2%	0.174	75.3%	65.4%	0.036
16	52.1%	66.4%	0.035	75.3%	66.9%	0.065
17	56.3%	68.2%	0.066	78.7%	68.0%	0.023
18	60.4%	66.8%	0.230	80.9%	69.4%	0.013
19	60.4%	66.6%	0.238	82.0%	70.6%	0.013
20	58.3%	68.4%	0.105	80.9%	71.7%	0.038

Table 76. Experimental vs. simulated decisiveness and Fisher's Exact test p-value. Low p-values indicate the two distributions are statistically significantly different.

Stage	No Feedback			Feedback		
	Experiment	Simulation	p value	Experiment	Simulation	p value
1	50.0%	51.9%	0.328	46.9%	50.2%	0.005
2	57.1%	58.8%	0.294	55.9%	52.7%	0.468
3	57.9%	63.7%	0.136	65.1%	54.1%	0.331
4	70.6%	68.2%	0.049	69.0%	57.6%	0.380
5	63.2%	71.5%	0.115	60.8%	59.7%	0.333
6	68.0%	74.2%	0.472	64.9%	60.9%	0.081
7	65.5%	78.1%	0.203	69.5%	62.0%	0.068
8	77.8%	81.7%	0.475	69.4%	64.4%	0.027
9	75.0%	82.3%	0.408	75.0%	66.9%	0.414
10	73.3%	83.8%	0.214	74.1%	68.2%	0.374
11	71.4%	85.4%	0.522	73.8%	69.7%	0.122
12	71.4%	87.8%	0.352	72.6%	72.1%	0.093
13	81.5%	88.2%	0.229	74.6%	73.9%	0.074
14	82.1%	89.8%	0.312	75.0%	75.8%	0.074
15	92.6%	90.3%	0.174	76.1%	76.3%	0.036
16	100.0%	90.7%	0.035	79.1%	78.6%	0.065
17	100.0%	90.9%	0.066	78.6%	78.8%	0.023
18	100.0%	92.2%	0.230	77.8%	80.1%	0.013
19	100.0%	94.9%	0.238	76.7%	81.0%	0.013
20	100.0%	95.0%	0.105	77.8%	81.5%	0.038

Table 77. . Experimental vs. simulated correctness and Fisher's Exact test p-value Low p-values indicate the two distributions are statistically significantly different.

Figure 30 compares the decisiveness and correctness after stage 20 for both the simulator and the experiment. This shows the simulator correctly captures the tradeoff between decisiveness and correctness seen in the experimental data; however, without as pronounced an effect as seen in the experiment.

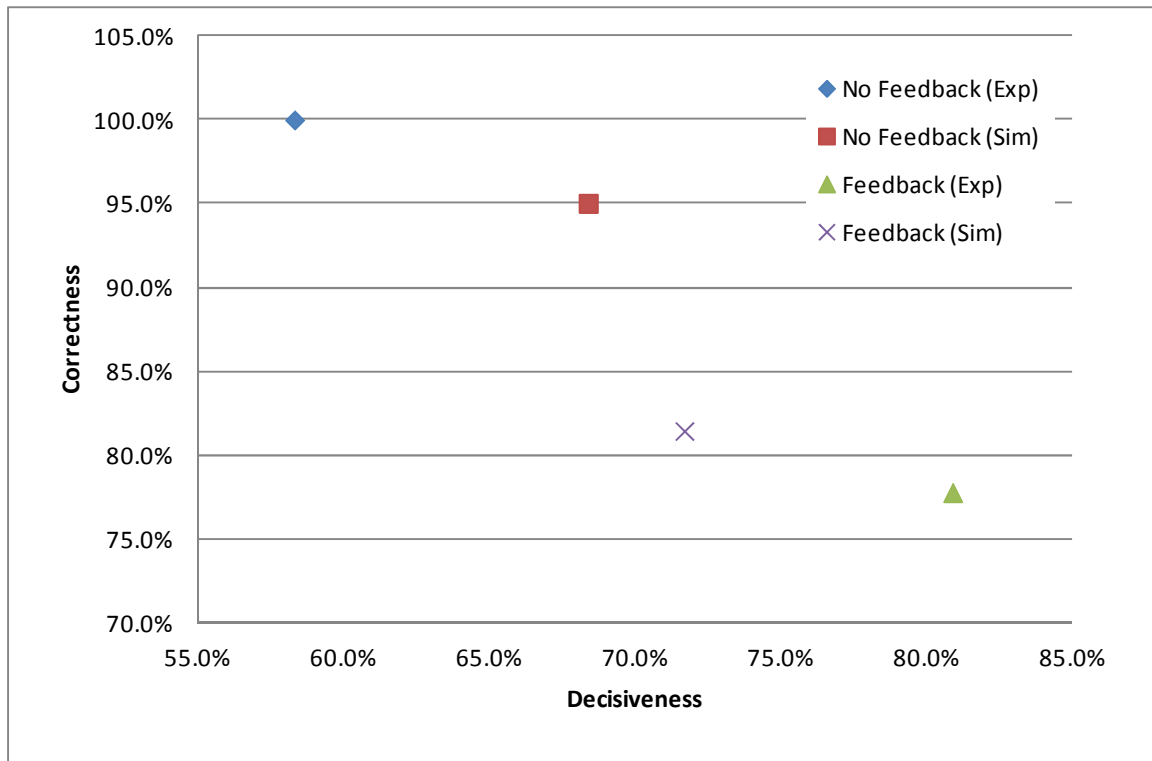


Figure 30. Comparison of experimental and simulated decisiveness and correctness after stage 20

5.2.5 Initial Simulation Results

To illustrate the application of a market simulator the prototype simulator was used to study the impact of changing the proportions of allocation styles on market performance. One hypothesis is that allocation strategies with low computation and low information usage (Figure 10) degrade overall market performance. This hypothesis was tested with the prototype simulator. No-feedback and frequency feedback markets were simulated with varying proportions of allocation styles. Table 78 shows the conditions simulated. First the proportions of strategies as derived in the random draws experiment were simulated as a baseline. Then a light and heavy computation mix were simulated. In the no-feedback market a light computation mix put more MaxiMin into the market at the expense of the DrawPct and DrawEV strategies. For the heavy computation mix,

the opposite was used with a smaller percentage of MaxiMin and larger percentage of DrawPct and DrawEV. Similarly, a light and heavy compute ratio of strategies for the frequency feedback markets was simulated. The light compute scenarios increased the proportion of MaxiMin and Mimic at the expense of the other, higher computation strategies. The heavy compute scenario reduced MaxiMin and Mimic to add more weight to the higher computation strategies.

	No Feedback Markets			Frequency Feedback Markets		
	As	Light	Heavy	As	Light	Heavy
	Expt.	Compute	Compute	Expt.	Compute	Compute
MaxiMin	33.9%	66.5%	16.5%	23.0%	40.0%	13.9%
DrawPct	48.1%	20.0%	60.0%	24.2%	6.1%	25.0%
DrawEV	9.5%	5.0%	15.0%	10.7%	1.8%	15.0%
Mimic	-	-	-	22.1%	40.0%	8.0%
PublicEV	-	-	-	1.0%	0.0%	0.0%
AllPct	-	-	-	6.6%	2.0%	10.0%
AllPctEV	-	-	-	0.0%	-	5.0%
AllDraws	-	-	-	4.2%	2.0%	10.0%
AllDrawsEV	-	-	-	0.0%	-	5.0%
No Fit	8.5%	8.5%	8.5%	8.1%	8.1%	8.1%
Markets Simulated	100,000	100,000	100,000	100,000	100,000	100,000
Decisiveness	67.5%	41.5%	78.5%	74.1%	35.5%	85.9%
Correctness	93.5%	81.0%	95.5%	80.6%	62.9%	84.6%

Table 78. Simulated market performance under different proportions of allocation styles.

Simulating 100,00 markets affirms the hypothesis. In both market types the light compute scenario reduces both decisiveness and correctness while the heavy compute scenario increase both decisiveness and correctness.

This illustration presents one example of the many types of analyses which a simulator can permit.

5.3 Considerations for Prediction Market Designers

The relevance of these results to the designer of prediction markets is explored.

5.3.1 Choosing a Feedback Type

Choosing the type of feedback to use in a prediction market is an important design consideration. Figure 31 shows the decisiveness and error rate as measured from the random draws experiment. Error rate is defined as 1-correctness and so is a simple remapping of Figure 13. However, this chart resembles an efficient horizon chart used in finance which is usually plotted as return on the y axis and risk on the x axis. So decisiveness is analogous to return in that the market designer wishes to get some information out of the market and error rate is analogous to financial risk.

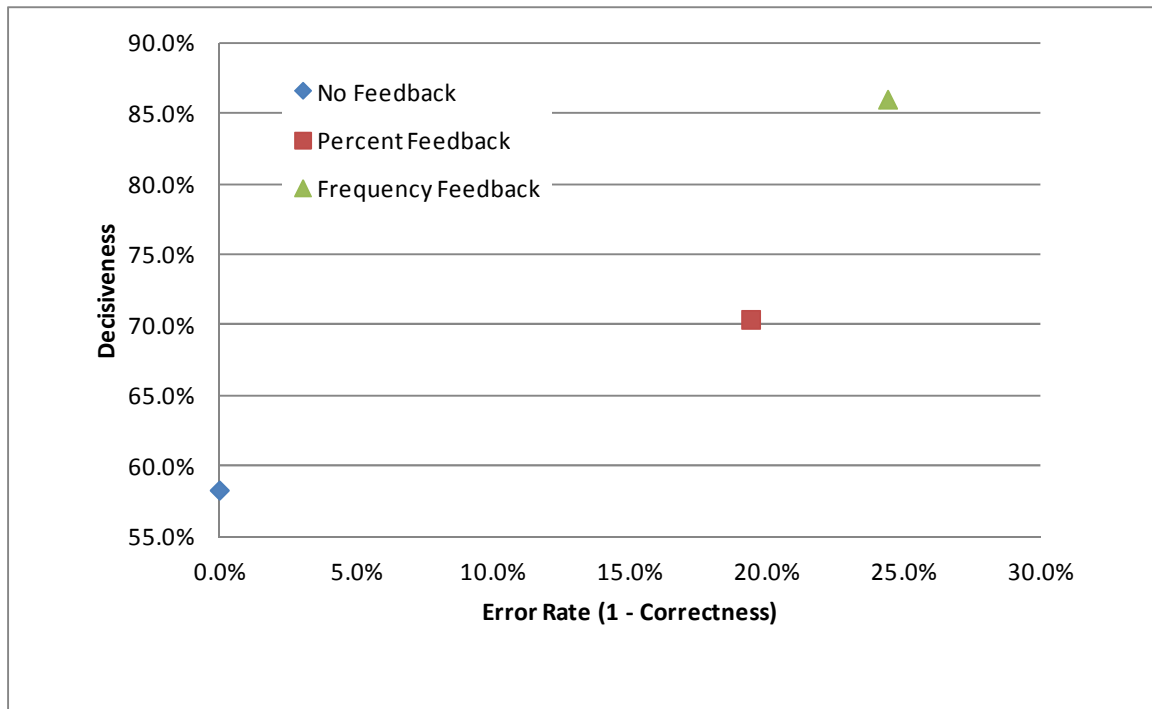


Figure 31. Decisiveness and error rate observed in the random draws experiment

So it is clear the market designer can trade some error rate to get more information from the market. The designer can judge the impact of an error in the market and choose accordingly. But, how much real value is the extra decisiveness of a feedback market really worth? Table 79 gives a summary of the no-feedback and feedback markets after stage 20 from the random draws experiment. Here, the results of Table 31 and Table 32 are shown as percentages. Feedback markets are providing significant results in 22.6% more markets than no-feedback markets. However, the incremental rates of judging incorrect values is 18% while the incremental correct values are only 4.6%. This can be considered that the correctness on the additional 22.6% market which were determined to be significant is only 20% ! (that's $4.6/(4.6+18)$).

	Correct	Incorrect	Not Sig.
No Feedback	58.3%	0.0%	41.7%
Feedback	62.9%	18.0%	19.1%
Change	-4.6%	-18.0%	22.6%

Table 79. Comparison of no-feedback and feedback markets.

As a different way to analyze the results, the product of correctness and decisiveness could be examined as a single metric of market performance.

$$\begin{aligned}
 & \text{Correctness} \times \text{Performance} = \\
 & = \frac{\# \text{ Correct}}{\# \text{ Correct} + \# \text{ Incorrect}} * \frac{\# \text{ Correct} + \# \text{ Incorrect}}{\text{Total}} = \frac{\# \text{ Correct}}{\text{Total}}
 \end{aligned}$$

This metric, percent correct, is shown in Table 80. By this metric, frequency feedback markets are the best performing.

Stage	Correctness x Decisiveness		
	No FB	Pct	Freq
1	16.7%	20.9%	13.0%
2	25.0%	23.3%	19.6%
3	22.9%	32.6%	30.4%
4	25.0%	37.2%	28.3%
5	25.0%	32.6%	37.0%
6	35.4%	41.9%	41.3%
7	39.6%	46.5%	45.7%
8	43.8%	48.8%	47.8%
9	43.8%	44.2%	50.0%
10	45.8%	48.8%	47.8%
11	41.7%	48.8%	52.2%
12	41.7%	48.8%	52.2%
13	45.8%	48.8%	56.5%
14	47.9%	48.8%	58.7%
15	52.1%	51.2%	63.0%
16	52.1%	51.2%	67.4%
17	56.3%	55.8%	67.4%
18	60.4%	58.1%	67.4%
19	60.4%	58.1%	67.4%
20	58.3%	58.1%	67.4%

Table 80. Percent correct by market type. Percent correct is the product of correctness and decisiveness. Results are shown by market stage (number of participants) for the random draws experiment.

The characteristics of the application may help the prediction market designer choose a market structure. Table 81 illustrates one selection framework. The expected signal strength and project characteristics are used here as the selection criteria. Project characteristics shown are Mission Critical and Exploratory are illustrated below:

- Mission Critical: e.g., prediction market is being used to support a forecast which will appear in the company's annual report
- Exploratory: e.g., prediction market is used as part of a marketing group's evaluation of new product concepts

The signal strength may be the designer's best guess or perhaps previous prediction markets have provided some indication (e.g., many prior prediction markets resulted in

'Not Significant' results). Where the signal is expected to be strong and the project is critical then the designer may favor no-feedback markets. When the signal is expected to be weak and the results are not mission critical then the designer may favor a feedback market.

		Project Characteristics	
		Mission Critical	Exploratory
Signal Strength	Strong	Favor Correctness - No-Feedback	Need both Correctness and Decisiveness, but favor Decisiveness - Feedback or Hybrid?
	Weak	Need both Correctness and Decisiveness, but favor Correctness - No-Feedback Hybrid?	Favor Decisiveness - Feedback

Table 81. Potential selection criteria for market structure.

The way the market results are to be used might also influence the structure to be used.

As discussed in section 2.2.3.2.2, the Ellsberg Paradox showed that humans are averse to ambiguity. If the market results are needed to convince an audience (e.g., senior management) then more Decisive and lower entropy feedback markets may be valuable.

5.3.2 Market Size (Number of Participants)

The number of participants is another parameter which may be under the control of the market organizer. Should the market be kept to a small, select group in the organization or made more generally available? It is expected that this would be heavily dependent on the signal strength (availability of private information) and the allocation styles used by the participants. The experimental data can be examined for some guidance. Figure 32 shows the decisiveness and correctness results for the no-feedback and feedback

markets by market stage. No-feedback market decisiveness is quite stable after 7 participants. However, the correctness continues to rise until it reaches 100% at stage 16. By contrast, decisiveness for feedback markets appears to keep rising over most of the experiment with a plateau being reached quite late in the experiment. Feedback market correctness appears to reach a maximum after approximately 15 participants.

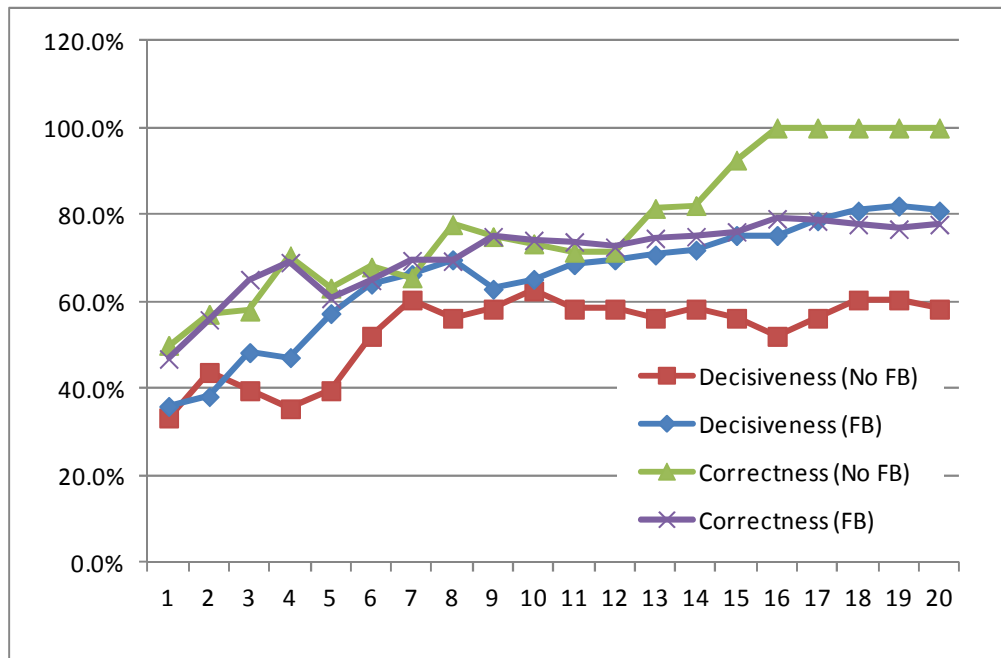


Figure 32. Decisiveness and correctness by stage (random draws experiment)

5.3.3 Guiding Participants to Improve Performance

Section 5.2.5 suggested that if the participants used more available information and more complex algorithms in the allocation of their tickets, the performance of the market would improve. Generally, these changes also improve the participant's financial performance as they approach the maximum expected value, and thus represent a win-win for both the participant and market organizer. Plott et al. (2003) gave participant's a review of Bayesian statistics and showed them a version of Table 22. A prediction

market organizers may use some type of similar information to improve the performance of their participants.

5.3.4 Screening

The random draws experiment included two types of participant screening while the mirage prone data experiment had no screening. It's possible that the participants in the mirage prone experiment who would have been screened out if screens were present may have had a significant contribution to the poor decisiveness and correctness performance seen in those results.

Table 26, shown earlier, showed the results from a no-feedback market in the mirage prone. Participant 4 received a draw of DDE and put all 100 of their tickets on state C. These 100 tickets represented nearly 40% of the tickets allocated to state C through the first 12 stages and created a market which did not display a significant mode. So, not only was private information excluded from the market, the spurious input adversely impacted the result. In feedback markets the result of a spurious impact can be even more detrimental, especially if that input occurs early in the market and creates an information mirage.

Two screening methods were tested and described in section 4.1.4 including an attention test and a screen on a no-feedback market. No-feedback markets have an inherent advantage as screening need not be done in real time. Feedback markets would require the screening to be complete before the participant were allowed to enter the market.

5.3.5 Synchronous or Asynchronous Markets

Most of the prediction markets being implemented today are synchronous. An asynchronous market has been shown in the present study. While this implementation was used primarily due to the design of the experiment, it provides a potentially useful type of market. For a no-feedback market design there is no reason to have a synchronous market. For a feedback market one might argue that those at the end have an advantage over those at the beginning. However, if the market is repeated on a regular basis (e.g., a monthly forecast) then participants could have different positions on each cycle. Additionally, synchronous markets place a greater burden on the implementation of the market, especially in terms of the software needed. And, synchronous markets place a burden upon the participants as they must be available at the specific time (and perhaps place) of the prediction market. If the market includes participants across various time zones, this can be logistically difficult. The present research suggest practitioners might consider an asynchronous approach.

5.3.6 Evaluating Prediction Market Results

The present study has defined a method to judge the statistical significance of results in terms of correct, incorrect or not statistically significant as discussed in section 3.4.2. This can be valuable for the prediction market organizer in the interpretation of their results. Still qualitative analysis of distribution shape of real world prediction market results may give clues to participants' insights. For example, dual peaks may give clues to dissonance being experienced by participants (Hopman, 2008).

5.4 Future Research

Areas of potential future research are now discussed.

5.4.1 Simulation

Further simulation development, testing and applications are here indicated. Several possible simulation projects were described in section 5.1. As mentioned, simulation can help to understand market conditions that may be difficult or time consuming to test empirically. Simulation work may be used to narrow down a list of ideas for an empirical study. Or, simulation could be used after an empirical study to gain insight into a phenomenon observed in the experiment.

Within a given market structure, there are several important characteristics which could impact performance. Some of these may be under the control of the organizer. Others may not be under the organizer's control, but an understanding of them may influence the organizer's choices.

The performance of the market as a function of the number of participants in the market is of great interest to the market designer (section 5.3.2). Depending on the circumstances it is likely under some degree of control to the designer. In some cases, the designer may determine that relatively few participants should participate and in this situation the designer would be keenly interested in the expected performance of such market.

Other factors of interest may be difficult to control but the designer might have some influence on the effect. For example, the designer may influence diversity by who is invited to participate in the market. Training for participants on cascade tendencies or maximizing expected payoffs might influence feedback effects. The impact of these changes could be estimated and simulated to study the impact on market performance.

Other factors may be outside the control of the organizer but the designer may have some understanding of these parameters which could help in the analysis of the market results. Key factors in this category are signal strength and the presence of noise. Signal strength relates to the amount of information available to market participants in the problem space. Noise relates to how clearly available information is interpreted. To illustrate, consider a forecasting prediction market. If the participants have a lot of information relevant to the forecast (e.g., inventory levels, customer insights, ecosystem factors, etc.) they are able to make more knowledgeable assessments. If the information is unclear (e.g., customers giving conflicting signals, channel partners giving ambiguous responses, etc.) or noisy then individual assessments are more difficult. Simulation studies could help to increase understanding of how the markets perform under varying levels of signal strength and noise.

Table 82 briefly describes some hypotheses on how key factors may impact correctness and decisiveness in prediction markets. These represent a sampling of ideas for a simulation research project.

Factor	Hypothesized Impact on Correctness	Hypothesized Impact on Decisiveness	Hypothesized Net impact
Number of participants	More participants increase correctness	More participants increase decisiveness	More participants improves performance
Diversity	Greater diversity increases correctness	Greater diversity increases decisiveness	Greater diversity improves performance
Degree of positive feedback	More positive feedback decreases correctness	More positive feedback increases decisiveness	Positive feedback involves a tradeoff
Degree of negative feedback	More negative feedback increases correctness in a system which already has positive feedback (by reducing information mirages)	More negative feedback reduces decisiveness correctness in a system which already has positive feedback	Negative feedback involves a tradeoff
Signal Strength	Stronger signal increases correctness	Stronger signal increases decisiveness	Stronger signal improves performance
Noise Levels	More noise reduces correctness	More noise reduces decisiveness	More noise degrades performance

Table 82. Hypothesized impact of key factors

5.4.2 Parimutuel Markets

Parimutuel markets were briefly mentioned in section 2.1.1.2. These were explored in detail by Plott et al. (2003). Table 83 summarizes the pros and cons of a parimutuel system.

Parimutuel systems have some interesting benefits. First, compared to a fixed winning system, they provide a measure of negative feedback. Since the market prize is split amongst winning tickets, as more tickets are allocated to a particular state the incentive to place more tickets on that state decreases. Another state, which may have a lower

probability of being correct, may offer a better expected value. This negative feedback should help in reducing information mirages.

A long shot bias has been observed in gambling (Woodland & B.M., 1994) as well as parimutuel market (Hurley & McDonough, 1995). In this situation, participants tend to overvalue long shots and undervalue favorites. This phenomenon was also observed in the experimental parimutuel prediction markets (Plott et al., 2003).

However, there are some clear negative issues with parimutuel markets. First, since the pool will be split amongst all winning tickets, there is incentive for participants to attempt to influence other members as they place their bets. For example, bluffing may occur. Second, if there is some indication of the accumulated bets, there is incentive to delay making one's bet until the last possible moment – when all the diverse information of the participants has been accumulated in the system. This is in contrast to a double auction process where participants are incentivized to bet early, when market inefficiencies may be present (and prices are low).

Pros	Cons
Provides a source of negative feedback - could provide a counter force to information mirages	Could foster disruptive strategic behavior – e.g., bluffing
May provide additional incentive to participants as they perceive large potential winnings	Incentivizes delaying the allocation of tickets (as group information is collected) – vs. double-auction which incentivizes quick action
	May cost the market organizer more in terms of payments since the entire pool of incentive prizes is paid each time - independent of the performance of the market

Table 83. Pros/Cons of parimutuel system

Comparing the parimutuel and a fixed winning system offers interesting tradeoffs between the payout the organizer must make and the perceived possible winnings by participants. In the parimutuel system the entire market prize is awarded each time, regardless of market performance (with the exception when no tickets are placed on the correct state). By contrast, the payout in a fixed winning system is unpredictable. It is bounded by the winnings per ticket multiplied by the number of tickets, but would rarely if ever reach that level. To the organizer's benefit, the payout is proportional to the performance of the market. The tradeoff here concerns the incentives perceived by the participants. In a fixed winning system the participant knows their maximum possible winning (the amount one winning ticket will fetch times the number of tickets). By contrast, in a parimutuel system there is the chance that one single ticket could win the entire market prize (if that one ticket is the only ticket placed on the correct state). To illustrate this imagine a corporate prediction market designer who has a budget of \$100. They want to run a market with 10 participants and are considering a fixed winning vs. a parimutuel system. How should they set up the prize structure? For the parimutuel they

may simply say the prize is \$100 and each of you get 100 tickets to allocate. Any participant has the opportunity (perhaps very small) of winning the entire \$100. That may provide a significant amount of incentive and the participants may expend considerable energy in making their selections. The fixed winning system is much more difficult to plan. If the budget is fixed they may indicate that each participant gets 100 tickets and each correct ticket is worth 10 cents ($\$100/1000$) budget compliance. However, in this case, each participant can win a maximum of only \$10. This may result in less mental energy being expended during the allocation process. Another option open to the organizer is to make tickets worth more than 10 cents and assume some maximum percentage of the tickets will be allocated on the correct state. That would provide more incentive to the participants but the organizer runs the risk of going over budget. In fact, this characteristic makes the design of a research experiment to compare fixed winning and parimutuel markets very challenging.

5.4.3 Hybrid Markets

Roust and Plott (1999) tested a two stage parimutuel market where the first stage had no feedback and the second stage included feedback. Other types of hybrid structures would be interesting to explore. For example, a hybrid market could be constructed where some members of the market receive feedback and others do not. A hypothesis is that a hybrid market's performance would lie somewhere on the curve formed by the pure markets of Figure 13.

5.4.4 Treatments to Improve Individual's Performance

As discussed in sections 5.3.3 and Plott et al. (2003) there may be opportunities to improve the way individuals allocate tickets which would improve the overall performance of the markets (as well as increase the individual's winnings). An experiment where one group received a certain treatment (such as being shown Table 22, the Bayesian probabilities given the different types of draw sets) compared to a control group should be able to test this hypothesis.

5.4.5 Correlation of Allocation Styles to Psychological Characteristics

Further analysis and empirical experimentation could help understand the relationship between allocation styles and individual's characteristics. A deeper understanding here could help to tailor the treatments as described in section 5.3.3.

5.4.6 Impact of Final Digit Bias

Figure 19 showed a bias by respondents to round allocations to multiples of 5 or 10. It would be valuable to understand if, and how, this phenomenon impacts the performance of the markets.

5.5 Summary

Prediction markets are emerging as an interesting tool for organizational knowledge management. Their ability to assimilate diverse information in the organization and bypass some of the human biases present in group processes is seen as a valuable instrument to improve business performance. The primary goal of the present study was

to gain insights into the operation and optimization of prediction markets. Specifically, the research sought to understand how these markets assimilate information and how this process is influenced by the presence, and type, of feedback.

The research supported the hypothesis that information assimilation in feedback markets is composed of two mechanisms - information collection and aggregation. These are defined as:

- Collection - The compilation of dispersed information - individuals using their own private information make judgments and act accordingly in the market.
- Aggregation - The market's judgment on the implications of this gathered information - an inductive process. This effect comes from participants integrating public information with their private information in their decision process.

Information collection was studied in isolation in no feedback markets and the hypothesis that markets outperform the average of their participants was supported. The hypothesis that with the addition of feedback, the process of aggregation would be present was also supported. Aggregation was shown to create agreement in markets (as measured by entropy) and drive market results closer to correct values (the known probabilities).

Information mirages were hypothesized as a key failure mechanism and their fingerprint was detected in the research data. It was suggested that the very mechanism of market information aggregation, the inductive portion of information assimilation, can inherently produce bubbles based on randomness in the processes. This illustrates that market bubbles in the form of information mirages can come about as the result of statistical

variations in private data coupled with *rational* judgment and decision processes; bubbles need not come from simple irrationality.

A primary hypothesis of the present study was that changing the presence and type of feedback supplied to the market can impact market performance because of its impact on this inductive process. The research showed that the presence of feedback increased the precision of the market at the expense of accuracy. The data also suggested that the type of feedback may modulate this process which may allow a prediction market organizer to tailor the market to the specific requirements of their task.

The way individual participants use information to make allocations was studied. In feedback markets the fit of participant's responses to various decision models demonstrated great variety. The decision models ranged from little use of information (e.g., MaxiMin), use of only private information (e.g., allocation in proportion to probabilities), use of only public information (e.g., allocating in proportion to public distributions) and integration of public and private information. Analysis of all feedback market responses using multivariate regression also supported the hypothesis that public and private information was being integrated by some participants. The subtle information integration results are in contrast to the distinct differences seen in markets with varying levels of feedback. This illustrates that the differences in market performance with feedback are an emergent phenomenon (i.e., one that could not be predicted by analyzing the behavior of individuals in different market situations).

The characterization of individual participant's use of information can lead to the development of models (and simulators) to understand the performance of different market structures. This might be used to predict potentially interesting performance characteristics of a prediction market under certain conditions that could later be tested empirically; such as a treatment that changes the way individuals use their information to allocate tickets and thereby improves the overall performance of the prediction market. Alternately, the characterization of individuals responses may be used with a simulator to explain an effect seen in a prediction market that had not previously been understood - an application of generative science to explain an emergent effect.

The results of this study have increased our collective knowledge of market operation and have revealed methods that organizations can use in the construction and analysis of prediction markets. In some situations markets without feedback may be a preferred option. The research has studied information aggregation and shown support for the hypothesis that it can be simultaneously responsible for the beneficial information processing in feedback markets as well as the harmful information mirage induced bubbles. In fact, a market subject to a mirage prone data resembles a Prisoner's Dilemma where individual rationality results in collective irrationality.

6 References

- Abramowicz, M. (2007). *Predictocracy: Market Mechanisms for Public and Private Decision Making*: Yale University Press.
- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica* 21, 503-546.
- Anderson, L. R., & Holt, C. A. (1997). Information Cascades in the Laboratory. *American Economic Review*, 87(5), 847-862.
- Anderson, N. H. (1981). *Foundations of Information Integration Theory*: Academic Press.
- Anderson, P. W., Arrow, K. J., & Pines, D. (Eds.). (1988). *The Economy as an Evolving Complex System*: Westview Press.
- Arthur, W. B., Durlauf, S. N., & Lane, D. (Eds.). (1997). *The Economy As An Evolving Complex System II*: Addison-Wesley.
- Arthur, W. B., Holland, J., LeBaron, B., Palmer, R., & Tayler, P. (1997). Asset pricing under endogenous expectations in an artificial stock market. In W. B. Arthur, S. Durlauf & D. Lane (Eds.), *The Economy as an Evolving Complex System II* (pp. 15-44): Addison-Wesley.
- Axelrod, B., Kulick, B., Plott, C. R., & Roust, K. A. (2007). Design Improved Parimutuel-Type Information Aggregation Mechanisms: Inaccuracies and the Long-Shot Bias as a Disequilibrium Phenomena: California Institute of Technology.
- Bachelier, L. (1900). *The Theory of Speculation*. Sorbonne.
- Baird, B. F. (1989). *Managerial Decisions Under Uncertainty: An Introduction To The Analysis of Decision Making*: John Wiley & Sons, Inc.
- Banerjee, A. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, 107(3), 797-817.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44, 211-233.
- Beinhocker, E. D. (2006). *The Origin of Wealth: The Radical Remaking of Economics and What it Means for Business and Society*: Harvard Business School Press.
- Benkler, Y. (2006). *The Wealth of Networks - How Social Production Transforms Markets and Freedom*: Yale University Press.
- Berg, H. Prediction Markets at Microsoft: Microsoft.
- Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2008). Results from a Dozen Years of Election Futures Markets Research. In C. R. Plott & V. L. Smith (Eds.), *Handbook of Experimental Economics Results* (pp. 742-751): Elsevier.
- Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis. *Commentarti Academiae Scientiarum Imperialis Petropolitanae*, 5, 175-192.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5), 992-1026.
- Binmore, K. (1991). *Fun and Games: A Text on Game Theory* D.C. Heath.
- Birchler, U., & Butler, M. (2007). *Information Economics*: Routledge.

- Birnbaum, M. H. (2004). Human Research and Data Collection via the Internet. *Annual Review of Psychology*, 55, 803-832.
- Brunnermeier, M. K. (2001). *Asset Pricing under Asymmetric Information: Bubbles, Crashes, Technical Analysis and Herding*: Oxford University Press.
- Camerer, C. (1992). The Rationality of Prices and Volume in Experimental Markets. *Organizational Behavior and Human Decision Processes*, 51, 236-272.
- Camerer, C. (1995). Individual Decision Making. In J. H. Kagel & A. E. Roth (Eds.), *The Handbook of Experimental Economics* (pp. 587-683): Princeton University Press.
- Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*: Russel Sage Foundation/Princeton University Press.
- Camerer, C., Ho, T. H., & Chong, J. K. (2004). A Cognitive Hierarchy Model of Games. *Quarterly Journal of Economics*, 109(3), 861-898.
- Choo, C. W. (1998). *The Knowing Organization: How Organizations Use Information to Construct Meaning, Create Knowledge and Make Decisions*. New York: Oxford University Press.
- Ciborra, C., & Patriotta, G. (1998). Groupware and teamwork in R&D: limits to learning and innovation. *R&D Management*, 28(1), 43-52.
- Cohen, K. J., Maier, S. F., Schwartz, R. A., & Whitcomb, D. K. (1983). A simulation model of stock exchange trading. *Simulation* 41, 181-191.
- Condorcet, M. d. (1785). *Essay on the Application of Analysis to the Probability of Majority Decisions*.
- Cook, R. D. (1977). Detection of Influential Observations in Linear Regression. *Technometrics* (19), 15-18.
- Cowgill, B., Wolfers, J., & Zitzewitz, E. (2008). Using Prediction Markets to Track Information Flows: Evidence from Google. Retrieved from
- Davenport, T., & Prusak, L. (1998). *Working Knowledge*. Boston: Harvard Business School Press.
- Dougherty, M. E., Mynatt, C. R., Tweney, R. D., & Schaivo, M. D. (1979). Pseudodiagnosticity. *Acta Psychologica*, 43, 111-121.
- Eddy, D. (1988). Variations in physician practice: The role of uncertainty. In J. Dowie & A. S. Elstein (Eds.), *Professional Judgment: A reader in clinical decision making* (pp. 200-211): Cambridge University Press.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence in the illusion of validity. *Psychological Review*, 85, 395-416.
- El-Gamal, M. A., & Grether, D. M. (1995). Are People Bayesian? Uncovering Behavioral Strategies. *Journal of the American Statistical Association*, 90(432), 1137-1145.
- Ellsberg, D. (1961). Risk, Ambiguity, and the Savage Axioms. *Quarterly Journal of Economics*, 75 (4), 643-669.
- Epstein, J. M., & Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*: Brookings Institute Press, The MIT Press.
- Erev, I., & Roth, A. E. (1998). Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. *The American Economic Review*, 88(4), 848-881.
- Eyster, E., & Rabin, M. (2009). Rational and Naive Herding

- Fama, E. (1965). The Behavior of Stock Market Prices. *Journal of Business*, 38, 34-105.
- Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of deindividuation in a group. *Journal of Abnormal and Social Psychology*, 47, 382-389.
- Fischhoff, B., Lichtenstein, S., Slovic, P., Derby, S. L., & Keeney, R. L. (1981). *Acceptable Risk*: Cambridge University Press.
- Fisher, R. A. (1954). *Statistical Methods for Research Workers*: Oliver and Boyd.
- Forsyth, D. R. (1999). *Group Dynamics*: International Thompson Publishing.
- Forsyth, R., Nelson, F., Neumann, G., & Wright, J. (1992). Anatomy of an experimental political stock market. *American Economic Review*, 82, 1142-1161.
- Friedman, D., & Rust, J. (Eds.). (1993). *The Double Auction Market: Institutions, Theories and Evidence - Proceedings of the Workshop on Double Auction Markets Held June 1991 in Santa Fe, New Mexico*: Addison-Wesley Publishing Company.
- Gibbons, R. D. (1992). *Game Theory for Applied Economists*: Princeton University Press.
- Gigerenzer, G., & Hoffrage, U. (1995). How to Improve Bayesian Reasoning Without Instruction: Frequency Formats. *Psychological Review*, 102(4), 684-704.
- Gilbert, N., & Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Berkshire, England: Open University Press.
- Gilmour, D. (2003). How to Fix Knowledge Management. *Harvard Business Review*, 17-18.
- Graham, J. R. (1999). Herding among Investment Newsletters: Theory and Evidence. *The Journal Of Finance*, 54(1), 237-268.
- Grether, D. M., & Plott, C. R. (1979). Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review*, 69, 623-638.
- Grossman, S. J. (1981). An Introduction to the Theory of Rational Expectations Under Asymmetric Information. *Review of Economic Studies*, 48, 541-559.
- Grossman, S. J., & Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets. *The American Economic Review*, 70(3), 393-408.
- Gupta, B., Iyer, L., & Aronson, J. (2000). Knowledge Management: practices and challenges. *Industrial Mgmt & Data Systems*, 100(1), 17-21.
- Hahn, R. W., & Tetlock, P. C. (Eds.). (2006). *Information Markets: A New Way of Making Decisions*: The AEI Press.
- Harris, L. (2003). *Trading and Exchanges: Market Microstructure for Practitioners*: Oxford University Press.
- Harsanyi, J. C. (1967, 1968). Games with Incomplete Information Played by "Bayesian" Players, I-III. []. *Management Science*, 14(3, 5, 7), 159-182, 320-334, 486-502.
- Hastie, R., & Dawes, R. M. (2001). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*: Sage Publications.
- Hayek, F. A. (1945). The Use of Knowledge in Society. *American Economic Review*, XXXV(4), 519-530.
- Hirshleifer, D., & Teoh, S. H. (2003). Herd Behavior and Cascading in Capital Markets: a Review and Synthesis. *European Financial Management*, 9(1), 25-66.

- Hong, L., & Page, S. E. (2004). Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385-16389.
- Hopman, J. W. (2007). Using Forecast Markets to Manage Demand Risk. *Intel Technology Journal*, 11(2).
- Hopman, J. W. (2008). *Managing Bias in Demand Forecasting*. Paper presented at the Informs.
- Hurley, W., & McDonough, L. (1995). A Note on the Hayek Hypothesis and the Favorite-Longshot Bias in Parimutuel Betting. *American Economic Review*, 85(4), 949-955.
- Johnson, N. F., Jefferies, P., & Hui, P. M. (2003). *Financial Market Complexity*: Oxford University Press.
- Jolly, R., & Wakeland, W. (2008). *Using Agent Based Simulation and Game Theory Analysis to Study Information Sharing in Organizations – The InfoScape*. Paper presented at the 41st Hawaii International Conference on System Sciences, Hawaii.
- Kahneman, D., & Tversky, A. (1973). On the Psychology of Prediction. *Psychological Review*, 80, 237-251.
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, XLVII 263-291.
- Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money*: Harcourt Brace and Co.
- Ladha, K. (1992). The Condorcet Jury Theorem, Free Speech, and Correlated Votes. *American Journal of Political Science*, 36(3), 617-634.
- Le Bon, G. (1895). *The Crowd*: The Viking Press.
- LeBaron, B. (2000). Agent based computational finance: Suggested readings and early research. *Journal of Economic Dynamics and Control*, 24(5-7), 679-702.
- Levene, H. (1960). Robust tests for equality of variances. In H. H. e. a. Ingram Olkin (Ed.), *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*. (pp. 278–292): Stanford University Press.
- Liebowitz, J. (Ed.). (1999). *Knowledge Management Handbook*: CRC Press.
- Lin, L., & Kwok, L. (2006). Challenges to KM at Hewlett Packard China. *KM Review*, 9(1), 20-23.
- Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method : techniques and applications*: Addison-Wesley.
- Lloyd, J. N., & Kotz, S. (1977). *Urn Models and Their Applications: An Approach to Modern Discrete Probability Theory*: Wiley.
- Long, D. W. D., & Fahey, L. (2000). Diagnosing cultural barriers to knowledge management. *Academy of Management Executive*, 14(4), 113-127.
- Mackay, C. (1932). *Extraordinary Popular Delusions & the Madness of Crowds*: Wells, Vt., Fraser Pub. Co.
- Malkiel, B. G. (2003). *A Random Walk Down Wall Street*: Norton.
- Maloney, M. T., & Mulherin, J. H. (2003). The complexity of price discovery in an efficient market: the stock market reaction to the Challenger crash. *Journal of Corporate Finance* 9, 453- 479.

- Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18(1), 50-60.
- Mantegna, R. N., & Stanley, H. E. (2000). *An Introduction to Econophysics*: Cambridge University Press.
- Marsaglia, G., & Tsang, W. W. (2000). The Ziggurat Method for Generating Random Variables. *Journal of Statistical Software*, 5(8).
- Mason, W., & Suri, S. (2010). Conducting Behavioral Research on Amazon's Mechanical Turk *Social Science Research Network Working Paper Series*.
- McPhail, C. (1991). *The Myth of the Madding Crowd*: Aldine de Gruyter.
- Miller, R. M. (2002). *Experimental Economics: How We Can Build Better Financial Markets*: John Wiley & Sons.
- Mitroff, I., & Turoff, M. (1973). Technological Forecasting and Assessment: Science and/or Mythology. *Technological Forecasting and Social Change*, 5, 113-134.
- Muth, J. A. (1961). Rational Expectations and the Theory of Price Movement. *Econometrica*, 29(6), 315-335.
- Nissen, M. E. (2006). *Harnessing Knowledge Dynamics, Principled Organizational Knowing & Learning*. Hershey, PA: IRM Press.
- Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* Princeton University Press.
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5).
- Phillips, L. D., & Edwards, W. (1966). Conservatism in a Simple Probability Inference Task. *Journal of Experimental Psychology*, 72(3), 346-354.
- Plott, C. R., & Sunder, S. (1988). Rational Expectations and the Aggregation of Diverse Information in Laboratory Security Markets. *Econometrica*, 56(5), 1085-1118.
- Plott, C. R., Wit, J., & Yang, W. C. (2003). Parimutuel betting markets as informational aggregation devices: experimental results. *Economic Theory*, 22, 311-351.
- Plummer, T. (2003). *Forecasting Financial Markets - The Psychology of Successful Investing*: Kogan Page.
- Porter, D. P., & Smith, V. L. (1995). Futures Contracting and Dividend Uncertainty in Experimental Asset Markets. *The Journal of Business*, 68(4), 509-541.
- Postmes, T., & Spears, R. (1998). Deindividuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, 123, 238-259.
- Resnick, M. D. (1987). *Choices: An Introduction To Decision Theory*: University of Minnesota Press.
- Rheingold, H. (2002). *Smart Mobs: The Next Social Revolution*: Perseus Books Group.
- Rosser, J. B. (1991). *From catastrophe to chaos : a general theory of economic discontinuities*: Kluwer Academic.
- Roust, K. A., & Plott, C. R. (2005). *The Design and Testing of Information Aggregation Mechanisms: A Two Stage Parimutuel IAM*. Working Paper. California Institute of Technology.
- Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, 15, 353-375.

- Scharfstein, D. S., & Stein, J. C. (1990). Herd Behavior and Investment. *The American Economic Review*, 80(3), 465-479.
- Schutte, M., & Snyman, M. (2006). Knowledge flow elements within a context - a model. *South African Journal of Information Management*, 8(2).
- Sharp, D. (2003). Knowledge management today: challenges and opportunities. *Information Systems Management*, Spring, 32-37.
- Shelley, R. (1989). *The Lottery Encyclopedia*: Byron Publishing Services.
- Shiller, R. J. (2000). *Irrational Exuberance*: Princeton University Press.
- Shleifer, A. (2000). *Inefficient Markets: An introduction to Behavioral Finance*. New York: Oxford University Press.
- Sigmund, K., Fehr, E., & Nowak, M. A. (2002). The Economics of Fair Play. *Scientific American*, 286(1).
- Simon, H. (1976). *Administrative Behavior*. New York: The Free Press.
- Slovic, P., & Lichtenstein, S. (1968). Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology Monograph Supplement*, 78(3), 1-18.
- Slovic, P., & Lichtenstein, S. (Eds.). (2006). *The Construction of Preference*: Cambridge University Press.
- Smith, A. (1776). *The Wealth of Nations*.
- Smith, V. L., Suchanek, G. L., & Williams, A. W. (1988). Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets. *Econometrica* 56 (5), 1119-1151.
- Sornette, D. (2003). *Why Stock Markets Crash: Critical Events in Complex Systems*: Princeton University Press.
- Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin.
- Sunstein, C. R. (2006). *Infotopia: How Many Minds Produce Knowledge*: Oxford University Press, Inc.
- Surowiecki, J. (2005). *The Wisdom of Crowds*: Anchor Books.
- Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How Mass Collaboration Changes Everything*: Portfolio.
- Tiwana, A. (2002). *The Knowledge Management Toolkit*. Upper Saddle River, NJ: Pearson Education.
- Turner, R. H., & Killian, L. M. (1972). *Collective Behavior*: Prentice Hall.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1130.
- Tziralis, G., & Tsiopoulos, I. (2007). Prediction Markets: An Extended Literature Review. *The Journal of Prediction Markets*, 1, 75-91.
- Viscusi, W. K. (1985). A Bayesian Perspective on Biases in Risk Perception. *Econ. Letters*, 17(1), 59-62.
- Vives, X. (2008). *Information and Learning in Markets: The Impact of Market Microstructure*: Princeton University Press.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*: Princeton University Press.

- Woodland, L. M., & B.M., W. (1994). Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *Journal of Finance*.
- Wurtz, A. H. (1997). A universal upper bound on power functions *UNSW Discussion Paper 97/17*. Sydney, Australia: University of New South Wales
- Young, G. O. (2008). Prediction Markets: Wisdom Of The Crowd Comes To The Enterprise: Forrester Research.
- Zeeman, E. C. (1974). On the Unstable Behavior of Stock Exchanges. *Journal of Mathematical Economics*, 1, 39-49.
- Zimbardo, P. G. (Ed.). (1969). *The human choice: Individualation, reason and other versus deindividuation, impulse and chaos* (Vol. 17): University of Nebraska Press.
- Zwicz, M. (2011). [Personal Communication].

Appendix - Survey Details

1. IRB approved Consent Form:

You are invited to participate in a research study. This research is being conducted by Richard Jolly, a student in Portland State University's Systems Science Program and School of Business. The research is part of a PhD dissertation and is being supervised by Professor Wayne Wakeland.

If you decide to participate, you will be asked to partake in an on-line simulated market. You will be given some information about the market which is unique to you and you will see how other market participants have acted. You will then be asked to make an allocation of tickets which represents your best guess for the future state of the market. You will participate in a number of these market activities over the course of about 15-20 minutes. You will receive a payment after participating in the study. In addition, the knowledge which the study produces may help others in the future.

None of your contact information will be kept with the data – so there will be no way to link your responses to you.

Your participation is voluntary. You may also withdraw from this study at any time. However, if you do withdraw, you will not be eligible for the payment.

If you have concerns or problems about your participation in this study or your rights as a research subject, please contact the Human Subjects Research Review Committee, Office of Research and Sponsored Projects, 600 Unitus Bldg., Portland State University, (503) 725-4288 / 1-877-480-4400. If you have questions about the study itself, contact Richard Jolly at amtsurvey2010@yahoo.com.

Clicking the 'Yes' button indicates that you have read and understand the above information and agree to take part in this study. Please understand that you may withdraw your consent at any time, and that, by clicking, you are not waiving any legal claims, rights or remedies. Please print a copy of this page for your own records.

Would you like to participate?

- ☐ Yes, I have read the instructions and would like to participate
- ☐ No, I would not like to participate

2. Instructions for Incentivized Poll

Instructions:

Please read the instructions carefully so that you can properly complete the survey.

How you earn your bonus

You will be participating in an experiment in the economics of market decision making. Your bonus will be determined from your allocation of tickets in a set of simulated markets. There are six states in each market: A, B, C, D, E and F. One of the six will be randomly chosen as the correct state for any given market. You will be given 100 tickets for each market which you can distribute any way you wish across the six states. For example, let's say you distributed your 100 tickets as follows:

A	B	C	D	E	F
4	42	9	21	17	7

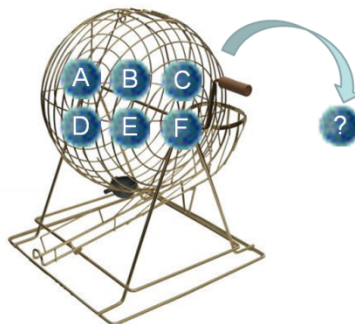
If the correct state were B, then, for this market, you would earn a bonus of 42% of the maximum bonus for that market.

You will be given information to help you distribute your tickets. But, first, we will explain how the correct state is determined.

The correct state

The computer will randomly determine the correct state for each market. The computer's process can be most easily understood by considering the analogy of drawing balls from a bingo cage.

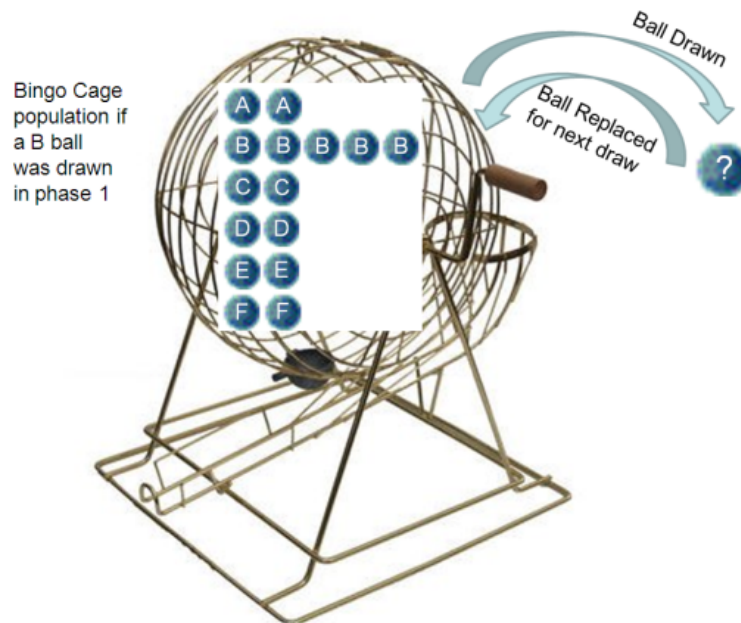
Six balls labeled A, B, C, D, E and F are placed in a bingo cage as shown in the figure below. One ball is drawn at random. The ball which was drawn becomes the *correct state* or the *solution* for the market. You will not be told the results of this draw – **it is what you are attempting to predict in the exercise.**



Phase 1 draw - the first bingo cage

Your Private Information:

You will also be given information which can help you distribute your tickets. This information is also randomly generated by the computer. Again, we use the bingo cage analogy to describe how this information is generated. Consider that balls will be drawn from a second bingo cage and these draws will be shown to you. This second cage will be **biased** based on the results of the draw from the first bingo cage. For example, let's say that the 'B' ball was drawn from the first bingo cage. The second bingo cage would then contain 5 'B' balls, 2 'A' balls, 2 'C' balls, 2 'D' balls, 2 'E' balls and 2 'F' balls as shown in the figure below. So, the odds that a B ball would be randomly drawn are higher.



The second bingo cage

The private data you will be shown is based on drawing balls from this second bingo cage. The exact method is called 'drawing with replacement.' A ball will be drawn from the cage, recorded and then replaced into the cage. The cage will then be spun and the process repeated until three draws are completed. The drawing with replacement process means that for each draw the contents of the bingo cage are the same. The results of these three draws will be reported to you on the screen in a graphic as below:

First Draw	Second Draw	Third Draw
B	F	F

This shows that the first ball drawn was a B ball. The second and third balls were both F balls.

The Markets:

The markets will be conducted online via this survey. You will be participating with others – spaced out in time. Between 10 and 20 people will participate in each market.

At the beginning of each exercise you will see the following graphic on your screen:

Tickets Remaining to Allocate	100
-------------------------------------	-----

State	A	B	C	D	E	F
Tickets Allocated	0	0	0	0	0	0

This indicates you have 100 tickets left to allocate across the six states – A through F.

Let's say, for example, that you then allocate 10 tickets for each state A through F by typing a 10 into the corresponding boxes in the table. The screen would now update to the following:

Tickets Remaining to Allocate	40
-------------------------------------	----

State	A	B	C	D	E	F
Tickets Allocated	10	10	10	10	10	10

This shows that 10 tickets have been allocated for each state and 40 are remaining to be distributed. You should continue allocating tickets until the 'Tickets Remaining to Allocate' indicates 0.

The screen you will see for each market will combine the private data - the draw of three balls from the bingo cage - and the ticket allocation graphic. The figure below is an example of the screen graphic.

Please show me the instructions again



Your Private Data:

First Draw	Second Draw	Third Draw
B	F	F

Tickets Remaining to Allocate	100
-------------------------------	-----

State	A	B	C	D	E	F
Tickets Allocated	0	0	0	0	0	0

Type your allocation into the shaded boxes

Do you understand the process and are you ready to begin?

- ☐ Yes
- ☐ No

3. Screen shot of Incentivized Poll Market

First Draw	Second Draw	Third Draw	
B	E	B	

[Please show me the instructions again](#)

Tickets Remaining to Allocate	30						
State	A	B	C	D	E	F	
Tickets Allocated	0	70	0	0	30	0	Type your allocation into the shaded boxes

4. Instructions for Feedback Markets

(note, this shows the instructions for frequency markets - the instructions were modified accordingly if the respondent was selected into percentage markets).

What others are doing:

In the final markets you will now have an additional source of information to use to make your ticket distributions - what others have done. In addition to your private draw, your market screen will also show the cumulative result of all those market participants who have preceded you.

The selections of those who have preceded you will be shown in a graphic as below:

The Market Data:

State	A	B	C	D	E	F
Tickets Allocated	44	122	19	22	44	149


This shows that 44 tickets have been allocated to state A, 122 to state B and so on. If you are the first participant in the market then the totals will be 0 for each state.

So, you now have two sources of information as to which ball is more likely to have been drawn for the market:

- the private information of your three ball draws from the bingo cage
- the cumulative allocations made by prior participants.

The screen you will see for each of these types of market exercises will be as below:

Please show me the instructions again



Your Private Data:

First Draw	Second Draw	Third Draw
B	F	F


The Market Data:

State	A	B	C	D	E	F
Tickets Allocated	44	122	19	22	44	149

Tickets Remaining to Allocate	100
-------------------------------	-----

State	A	B	C	D	E	F
Tickets Allocated	0	0	0	0	0	0

Fill in the shaded boxes with your allocation



Your Bonus in these markets:


You bonus will be calculated as mentioned before - based on the percentage of tickets you correctly allocate. There is no pooling of winnings - so how others allocate their tickets does not affect your bonus.

Do you understand the market information and are you ready to begin?

- ☐ Yes
- ☐ No

5. Screen Shot of Feedback Market

First Draw	Second Draw	Third Draw
C	E	C

[Please show me the instructions again](#) 

State	A	B	C	D	E	F
Tickets Allocated	0	0	0	0	0	0

Tickets Remaining to Allocate	100					
State	A	B	C	D	E	F
Tickets Allocated	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

Type your allocation into the shaded boxes

Only 3 market(s) left. Remember you must complete the *entire* survey to be eligible for payment.